AD-A038 597 BOLT BERANEK AND NEWMAN INC CAMBRIDGE MASS F/G 5/8
CRITICAL REVIEW AND ANALYSIS OF PERFORMANCE MODELS APPLICABLE T--ETC(U)
MAR 77 R W PEW, C E FEEHRER, S BARON F44620-76-C-0029
UNCLASSIFIED BBN-3446 AFOSR-TR-77-0520 NL

1 OF 4
ADA038597

4

597

Bolt Beranek and Newman Inc.

Report No. 3446

Critical Review and Analysis of Performance Models Applicable to Man-Machine Systems Evaluation

March 1977

Submitted to:
Air Force Office of Scientific Research

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

18

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 19 AFOSR - TR - 77 - 0520	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) 6 CRITICAL REVIEW AND ANALYSIS OF PERFORMANCE MODELS APPLICABLE TO MAN-MACHINE SYSTEMS EVALUATION.		5. TYPE OF REPORT & PERIOD COVERED Interim Scientific Report, 1 Oct 75 — 30 Sep 76
		6. PERFORMING ORG. REPORT NUMBER BBN Report No. 3446
7. AUTHOR(s) 10 Richard W. Pew, Sheldon Baron, Carl E. Feehrer, Duncan C. Miller		8. CONTRACT OR GRANT NUMBER(s) 15 No. F44620-76-C-0029 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS BOLT BERANEK AND NEWMAN INC. 50 Moulton Street Cambridge, MA 02138		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F/2313A4
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NL) Bolling AFB DC 20332		12. REPORT DATE 11 MARCH 1977
		13. NUMBER OF PAGES 300 12 305 p.
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 14 BBN-3446		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

16 2313 17 A4

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

man-machine
human performance
modelling and simulation
human reliability prediction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report focuses on the review of potentially relevant models and on the identification of issues in model development and application that may have an important impact on models for large-scale, man-machine systems. A detailed and critical evaluation of several classes of human-performance models is presented. Interrelations among existing models are examined and an evaluation is made of the needs and gaps in the technology. Modelling issues are identified and research recommendations indicated. Approximately forty models or model techniques that have some applicability to the simulation modelling program are abstracted in the Appendix.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

060100

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

JB

BBN REPORT NO. 3446

CRITICAL REVIEW AND ANALYSIS OF PERFORMANCE MODELS APPLICABLE TO MAN-MACHINE SYSTEMS EVALUATION

Richard W. Pew
Sheldon Baron
Carl E. Feehrer
Duncan C. Miller

MARCH 1977

Prepared under Contract F44620-76-C-0029

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
DIRECTORATE OF LIFE SCIENCES
BUILDING 410
BOLLING AIR FORCE BASE
WASHINGTON, D. C. 20332

PROGRAM MANAGER: Dr. Alfred R. Fregly

SYSTEMS RESEARCH BRANCH
HUMAN ENGINEERING DIVISION
AEROSPACE MEDICAL RESEARCH LABORATORY
AMRL/HEB
WRIGHT-PATTERSON AIR FORCE BASE
OHIO 45433

AIR FORCE TECHNICAL MONITOR: Dr. Robert Mills

ACKNOWLEDGMENT

The Air Force Technical Monitor for this work was Dr. Robert G. Mills. We wish to thank him and BBN personnel Dr. David Getty, Ms. Charlene Long, Ms. Ann Sarajian and Ms. Mary Tiernan for their contributions to the project.

TABLE OF CONTENTS

Table of Contents (cont.)

Table of Contents (cont.)

Executive Summary

This report describes the results and recommendations derived from an extensive survey of existing human performance models and modelling approaches applicable to the design and evaluation of large-scale command and control systems. The focus is on models derived from a control- and decision-theoretic framework, the modelling literature in human information processing, and the collection of models and data-bank formulations originally derived from the reliability and network-simulation literature.

The most successful modelling efforts seem to have grown out of situations where formal models of the task environment are well developed, such as in feedback control tasks, detection tasks, and well-defined probabilistic decision-making tasks. Further, in these areas, the most successful of these models arise when the researcher can express formal criteria of optimal performance as reflected in the optimal control formulation of manual control or the ideal observer in target detection and recognition tasks. These observations emphasize the importance to successful modelling efforts of being able to express the goals and success criteria used by human operators in formal quantitative terms. One difficulty for modelling behavior in more complex procedural tasks arises from the inherently multi-dimensional, multi-level, time-varying array of criteria and strategies that an operator applies in accomplishing these tasks.

It is interesting to note that the optimal control model and those information-processing models derived from a decision-theoretic framework are mutually compatible; this suggests the possibility of integrating and generalizing them to provide a single modelling framework that could be applied to vehicle

control, supervisory monitoring, surveillance, signal identification, and decision making, all of which are tasks of major interest in military system design and evaluation.

In the area of intellectual performance, however, modelling efforts have not produced practically useful results, either in areas where an explicit algorithm might be specified or with respect to general problem-solving performance, where a wide range of performance strategies are available. To represent these kinds of performance will require either structuring the problem so that results are not sensitive to differences in strategy or resorting to atheoretic representations derived from empirical measurements obtained in the specific task context.

In addition to the substantive reviews of modelling approaches, several methodological issues have been identified:

(1) The problem of validation of models of the scope considered here is a difficult one, requiring further research.

(2) Models exist at many different levels of specificity. A substantive issue concerns the identification of the level of specificity at which to define a model in relation to the goals of system performance prediction desired. Depending on the level of specificity that is appropriate, one must consider whether to take a top-down or a bottom-up approach. A top-down approach begins with goals of performance prediction and represents performance only down to level required to meet these goals. A bottom-up approach begins by defining the elemental components of human performance and synthesizes them into a model that predicts the desired aspects of performance.

(3) Bottom-up approaches to modelling continually face the issue of how to combine sub-task or task models into higher-level structures in such a way that the potential interactions resulting from their combination are accounted for. Additive combinations of component response times and multiplicative combinations of response accuracies are frequently not valid, and their applicability must be evaluated in each new synthesis.

(4) The current state of theory and understanding of human Performance is inadequate to represent many kinds of behavior observed in real task situations. The model developer is left with many arbitrary parameters that must be defined on the basis of observed performance. When the number of such free parameters approaches the number of performance measures to be predicted, the predictive power of the model is severely compromised.

On the basis of our review, recommendations are introduced for further research and development of large scale systems modelling efforts. These recommendations include:

(1) Development of a test-bed facility in which to evaluate alternative model formulations of common task environments and to conduct empirical validation studies to compare model predictions with actual human performance.

(2) Methodological research on (a) the implications of combining sub-task or information processing component models on system performance in the aggregate, (b) validation of large scale simulation models, and (c) development of guidelines for the acceptable number of free parameters in useful predictive models.

(3) Recommended further model development in topical areas of high priority for representation of command and control systems. Two such areas are supervisory control and monitoring and the prediction of team performance on the basis of performance of individual team members.

(4) Advancing the state-of-the-art with respect to the specific modelling approaches discussed in the body of the report.

From an overall perspective, we believe that integrative models of human performance compatible with the requirements for representing command and control system performance do not exist at the present time. What is available is a collection of component models and modelling principles formulated in a variety of frameworks, which might be drawn together to build an eclectic model for particular task situations of interest. On the basis of our present level of understanding, assembly of the components will call for substantial effort and is likely to require many assumptions about particular aspects of performance. If one is to have confidence in the product generated in this way, several iterative validation steps will be required.

1.0 INTRODUCTION

The Systems Research Branch of the Human Engineering Division of the Aerospace Medical Research Division (SRB) has undertaken a long-term program to develop and exploit simulation and modelling technology in the design and evaluation of large-scale systems. In support of this goal, BBN has initiated a three-year program of research to examine existing human performance models and approaches to modelling performance and to conduct studies exploring potential extensions of this technology.

Our goal is to develop the beginnings of a handbook-like document that would be useful to systems designers and analysts embarking on a systems modelling effort. Our research program assumes that computerized models of the system under consideration are to be constructed, and that these models will take into account the behavior of the human operators interacting with the system. By exercising these models, systems engineers wish to predict the effects of changes in system parameters before proceeding to full-scale simulation efforts and the operational evaluation of prototype systems.

In the first year of this program, we have focused on the review of potentially relevant models and on the identification of issues in model development and application that may have an important impact on the success of such modelling efforts. In the second and third years, we propose to conduct modelling studies in collaboration with personnel of SRB to explore quantitatively and in detail several key issues in model development.

This document represents the final report of the first year of effort.

1.1 Background

The current state of methodology for the development of large-scale, multi-man command and control systems, particularly tactical systems, is largely an art. The designers formulate goals and system objectives. They examine the available technology and produce a conceptual design that specifies the roles of hardware, software, and human resources. They then proceed to build either a manned simulation of the basic design or a prototype system that can serve as a test bed for evaluating the approach and for exploring alternative implementations of the basic approach. HED has pioneered in the application of the simulation approach for three different Air Force Systems, BUIC, AWACS, and Remotely Piloted Vehicle/Drone (RPV) Control, and has demonstrated the feasibility and cost effectiveness of preceding the prototype development phase by a simulation phase in which formal experiments of differing implementations may be undertaken.

HED is also looking ahead to the possibility of further efficiencies in system development through the use of full computer modelling of the hardware, software, and human performance components of new systems that would permit design trade-off studies to be undertaken in advance of full-scale manned simulation.

To this end they have developed the SAINT simulation language (Wortman, et.al., 1974), a dialect of GASP that has been specifically designed to make it possible to represent the adaptive capacities of human performance as well as to integrate discrete and continuous variable models in a single simulation language. A further feature of SAINT is the compact and explicit notation and software

operating system that permits easy translation from a flow chart representation of the system to a computer-coded model.

The basic approach to modelling represented in SAINT and other event-oriented modelling languages is to analyze the system flow chart into a series of processes related to each other through a contingently-branching structure. For each process, the necessary initial conditions for process activation are specified. Then a specified distribution of completion times and a probability of successful completion is assigned to the process. These parameters of each process may also be modified contingently while a simulation is in progress. Each time the model is exercised, a particular sample of each random variable is chosen and the contingent effects are assessed so that a unique trace through the multi-path flow chart is completed. Aggregated values for completion time and probability of completion are then derived from repeated runs of the simulation. The model is exercised sufficiently to accumulate reliable statistics for the major dependent variables under study, and it is these statistics that are the output of the model.

The usefulness of SAINT and of the modelling approach to system design have been explored in two major modelling studies. A model of a 3-man AWACS surveillance system was built using SAINT in parallel with the comparable manned simulation of the system (Mills, 1976). It was shown that a verified model could be obtained to describe system performance for the specific conditions under study. However, a great deal of effort and an estimated 50 iterations of the model form and parameters were required to debug it and to produce satisfactory matches to the simulation data. Further, it was shown that the model did not generalize very well to predict group performance when parameters representing the performance of individual operators acting in isolation were used.

In a second application, an upgraded version of the SAINT language was used to model the performance of the RPV/Drone Control System for which several manned simulation studies have been complete (Wortman, et.al., 1975). With a similar large effort to debug and tune the parameters of the model, a successful representation of the major system performance variables of RPV System Study II was obtained (Mills, Bachert, and Aume, 1975). Again, although generalization was not specifically tested, examination of the model suggests that it would not be particularly useful for design trade-off studies because the model was not effectively modularized to allow extension to new conditions without major adjustments in model structure and parameter settings. It was necessary for the model implementers to make too literal a representation of particular features of the situation in order to achieve the successful representation of overall performance. Said another way, the processes modelled were tasks that included components of both human and machine performance. Changing systems variables will produce changes that interact with human performance in new ways that are not easily separable in the model. Whether these difficulties are inherent in the modelling approach to describing large scale systems or are only limitations of the specific cases undertaken thus far remains to be seen. It does seem clear, however, that further progress in such modelling will require further development of modelling methodology and validation with specific attention to the representation of human performance in such systems in a way that has the potential for generalization to answer real systems questions. The models for human performance need to be structured to be compatible with the representation of the major systems variables of interest to the systems design engineer and not simply embedded implicitly in models for the major tasks involved.

Other laboratories (for example, Sandia, Naval Air Development Center, and Applied Psychological Services) have been developing large-scale modelling methodologies with varying degrees of similarity to the work in the Air Force. However, few (if any) of these efforts have had available resources equivalent to the HED multi-man simulation facility in which models may be explicitly compared and validated against the results of real-time, manned simulations. Thus the Air Force has a unique opportunity to advance the state-of-the-art in system modelling technology.

It is to this end that Bolt Beranek and Newman has undertaken this program of research. During the first year we have examined the rather extensive psychological and systems literature on models of human performance. Our sources have included:

Psychological Bulletin
Psychological Review
Journal of Experimental Psychology
Journal of Mathematical Psychology
Human Factors
IEEE Transactions
International Journal of Man-Machine Studies
Operations Research
Proceedings of Annual Conferences on Manual Control
JANAIR Reports
Technical Reports on Air Traffic Control
Books and Monographs on Information Processing Models
Miscellaneous Contractor and Government Reports

We have struggled repeatedly with questions concerning the scope of models appropriate to this review and evaluation. Models exist at many different levels of formality, ranging from verbal analytic statements of principles to closed-form mathematical solutions. The term model means something quite different to a social psychologist than to a physicist or engineering designer. One can precipitate extensive philosophical arguments by attempting

to distinguish a model from a theory. Rather than attempting a formal definition of a model, we have developed the following set of criteria for deciding whether to include a particular model in our review. Several of these criteria are pragmatic. Others relate to what we believe are the essential characteristics of quantitative predictive models.

1. Relevance. We are taking as our primary definition of relevance the kinds of activities required of operators in the RPV command and control system. We believe this system is representative of many of the characteristics of highly-interactive, large-scale, multi-operator systems in general and of a variety of tactical command and control systems in particular. We believe this focus is not overly restrictive of the more general domain of man-machine systems of interest. Specifically, we were seeking models for the following activities:

a) Target detection and identification.
b) Supervisory monitoring, display scanning, and surveillance.
c) Target tracking and manual control.
d) Alphanumeric data entry.
e) Graphical data entry.
f) Mental computation, estimation, and probability assessment.
g) Algorithmic problem solving.

2. Quantitative Predictions. Many models encountered in scientific journals provide a structure for a process or propose hypotheses about how an activity should be represented, but are not carried through to the point of making quantitative predictions about observable outputs or performance measures. Such models are of little value for our purposes.

3. Computational Complexity. We are excluding from consideration models whose computational complexity is so great as to preclude practical application to problems of interest.

4. Parameter Determination. In order for a model to be predictive, its parameters must be determinable, either *a priori* from other sources or from the constraints imposed by a particular application. A difficulty with many of the models available in the literature is that their parameters, if specified at all, have been determined only for a specific laboratory task and are unlikely to generalize to other contexts. While we have included several models subject to this criticism, we have tried to point out their limitations.

5. Reliability. A model is reliable, by definition, if repeated applications yield the same or substantially similar results. It should be noted, however, that it is sometimes difficult to judge the reliability of recent models, simply because many of them have not yet had the benefit of repeated application.

6. Validation. Tests of the validity of a model may range from informal demonstrations that it produces reasonable results to formal tests of the goodness of fit of the model's predictions to independent data. We are excluding from consideration proposed models for which no attempts at validation have been made. Formal tests of validity, however, are likely to exist only for models addressing individual tasks or components of human performance. For evaluating larger scale systems models, less formal evaluations may be the most we can expect.

Thus our goals have been to identify particular models that are sufficiently explicit and have promise for application to military command and control system modelling activities. In the course of this search we have also sought to identify issues and problems that limit model usefulness and that must be addressed if we are to achieve the broader goal of useful man-machine modelling technology.

1.2 Outline of Report

The remainder of this report is organized into four main sections. In Section 2 we have provided a detailed and critical evaluation of the classes of human performance models we have examined. This review is organized on the basis of various approaches to modelling rather than the contexts to which the models are applicable. We believe this will provide a more coherent presentation of the state-of-the-art. This section concludes with an examination of the interrelations among existing modelling efforts and an evaluation of the gaps between needs and available models. In Section 3 we discuss the issues and problems that we have identified in our review of models and modelling approaches. Section 4 describes recommendations for further research that is needed to advance our ability to use modelling as a system design tool. Finally, in Appendix A we provide abstracts of the models we have found that we believe have some measure of applicability in the system modelling context.

2.0 CRITICAL SURVEY AND REVIEW OF MODELS POTENTIALLY USEFUL FOR SIMULATION MODELLING

2.1 Data Bank Formulations

It became evident during efforts in the early 1950's to incorporate estimates of human reliability into estimates of total system effectiveness that relevant human performance data did not exist in useable form. The first attempt to extract such data from the literature of general and applied psychology and human factors, and to represent it in a tabular array useable by design engineers, was conducted by Payne and Altman at the American Institutes for Research (AIR) in 1962 (see Meister, 1964). Called "Data Store," this compilation describes the characteristics of various controls and displays, the minimum times required to operate specified controls, and the reliabilities with which the operations can be expected to be carried out.

Since Data Store, there have been repeated attempts to create data bases from which human performance times and error rates can be extracted and utilized. A number of these, though relatively complete with respect to specification, have not been implemented (see, for example, Meister and Mills, 1971; Verdi and Smith, 1966). Of those that have been implemented, some simply represent extensions of the list of elements contained in the original Data Store while others exhibit differences in basic organizational level of aggregation - that is, in the definition of what behavioral or equipment level to assume for tabulation of atomic events - or in task specificity.

None of the data bank formulations to date are models in the sense that we have used that term elsewhere in this report. They do, however, represent a chief source of data for modelling and simulation efforts and, as a result, have an important place in a review of the human performance modelling literature. In this section we shall discuss briefly some of the characteristics of data banks and attempt to highlight some issues surrounding their formulation and exploitation.

2.1.1 Performance Area of Concern

Almost all attempts to create data banks are limited to a specific area of concern and, because of the time, effort, and cost involved in developing and maintaining them, few seem to evolve beyond that original area. Data Store was originally conceived in support of work in display and control evaluation, and despite an interest in broadening its scope (see, for example, Altman 1967), the formulation remains most useful when employed in connection with the original purpose. The SHERB (Sandia Human Error Rate Bank) system was developed in the context of an interest in human performance in process-control operations. The HERALD data bank, developed by engineers at Lockheed-Georgia, is limited in interest to the interpretation and operation of displays and controls by an aircraft pilot and crew. Another example of this specificity is the formulation by Irwin, Levitz, and Freed (1964).

A possible exception to the generalization that data banks tend to be limited in their purview may be the system recently conceived by Blanchard (1973) for the Navy, which would, when completed, contain twelve or more smaller banks including the following:

1) Operational Performance Baseline Data
2) General Human Performance Capability Data
3) Psychophysiological Data
4) Personnel Cost Data
5) Human Locomation Data
6) Operational Requirements Information
7) Equipment Reliability Data
8) Micro-Behavior Data (eye movements, hand movement, etc.)
9) Maintenance Activity Data
10) Hardware Component and Specification Data
11) Shipboard Work Standards and Baseline Data
12) Human Engineering Standards and Specifications

2.1.2 Data Base Organization

Two basically different approaches have been taken to the design of data banks. One of these, exemplified by Data Store, places prime emphasis on the hardware item with respect to which the behavior to be predicted occurs. In this organization, a given hardware item is analyzed in terms of its components and a performance reliability estimate, established on the basis of a study of available literature or actual experimentation, is assigned for each component. A consequence of this approach is that the values on which the prediction of performance at a task level is to be based exist at a very molecular level in the data base and must be aggregated according to some appropriate rule in order to provide the desired prediction. Unfortunately, the nature of this "appropriate rule" is frequently in doubt.

An alternative to the molecular approach is illustrated by the SHERB organization. Here, the basic unit of analysis is the task itself and the procedural context in which that task occurs. Reliability data on specific types of errors with specific types of equipment are collected under such broad behavioral task descriptions as "assembly," "inspection," "installation," etc. This organization has at least two consequences: (1) it helps to obviate the aggregation of molecular reliabilities, at least at the level of single tasks; and (2) experimental and situational factors that may have been important determiners of the data employed to establish the reliability values, and that may or may not be present in the context of the tasks to be predicted, are more easily accumulated and assessed.

It is impossible to decide objectively at this point which of these two organizations is the better, since in no case of which we are aware has one been pitted against the other in the solution of the same performance prediction problem.* It may even prove to be impossible if such a confrontation were arranged, unless the latter also took into account the characteristics and training of different populations of users. It may be the case, for example, that the molecular-type organization is more "natural" and leads to more successful exploitation by the hardware and systems engineering communities and that the molar-type is more "natural" and exploitable by the human engineering community. To the extent that this is the case, it would seem appropriate to exploit current methods for storing data in minimally constrained computer file structures and for providing access by whatever interrogation conventions are felt by a given user to be most productive.

*There may be a number of reasons for this. One is the fact that, as we noted earlier, most data banks are compiled for the purpose of answering questions in a specific performance area and have little enough in common that a direct comparison would be inappropriate. A second is that the typical impetus for developing a new formulation is dissatisfaction with an old one. From one point of view, then, the decision as to which is superior has already been made by the time the design of the replacement is complete, and there may be little incentive to conduct a comparative evaluation.

2.1.3 Data Base Contents

All of the models for which current data banks are potentially useful are concerned with predicting the reliability with which the human operator will perform some assigned task or set of tasks. Some are also concerned with the time taken to perform the task. As a consequence, most data banks contain both reliability and completion time information, either in the form of point estimates (expected reliabilities, minimum or expected times), the observed distributions, or parametric data from which distributions can be derived.

Though there are few differences among designers and users regarding the most useful ways to represent temporal information, there is some controversy surrounding the nature of the most appropriate reliability statistic. As indicated in the summary, AIR Data Store contains, for each task element defined in the data base, a point estimate of the probability that that task element will be performed correctly, defined as unity minus the aggregated error probabilities obtained from the literature. Further, it contains, for each critical value or range of values for parameters associated with the equipment used, a point estimate of the probability that the parameter value will lead to successful performance. In this formulation, then, reliability is defined simply in terms of probable success without respect to time.

In some applications, the rate at which errors are produced is of critical concern, and there is a desire to include in the data base some indication of "failure rate" similar to that defined with respect to hardware operation. As Swain (1970) points out, such data tend to be difficult to come by since in most otherwise applicable field studies (e.g., in the area of quality assurance), the primary interest is in errors per unit time. The situation is much the same with respect to laboratory data, since most investigators report performance in terms of total number of errors per experiment or per experimental task.

2.1.4 Predictive Accuracy as a Function of Number of Parameters

All modeling techniques that begin by analyzing the performance of a given task into discrete activities of the sort represented in Data Store must include some rule for resynthesizing these molecular activities into a more molar entity, possibly the original task itself or an even higher-level aggregate such as a total procedure made up of n tasks. As we have noted, the rule used most often is the simple product rule, in which reliabilities associated with molecular activities are multiplied together to achieve an estimate at the desired level of aggregation.

It should be clear from our discussion that the reliability (and time) values that appear in human performance data banks represent approximations and "best estimates," and, as such, may be incorrect for any given application. One might expect, then, that if these values are aggregated with the product rule, the accuracy of the resulting estimate would somehow depend upon the number of task parameters the analyst has chosen to identify.

The only systematic study of this possibility uncovered in our review was made by Swain (1967). This investigator performed a Monte Carlo simulation on artificial (but meaningful) tasks constructed from AIR Data Store. Each iteration provided an estimate of the probability of one or more errors in performance. The resulting "prediction," arrived at by the rules established for use of Data Store (see Abstract No. 1) was then compared against the prediction of a nominal model that employed the average value of all relevant behavioral unit reliabilities raised to a power equal to the number of parameters considered:

$$Q = 1 - (\bar{p})^n$$

where

Q = the probability of one or more errors

$\bar{p}$ = the average value of behavioral unit reliabilities

n = the total number of parameters

Swain found that for a large number of parameters ($n = 40$), the nominal rule provided a very good approximation of the value derived from the simulation, but that for smaller numbers of parameters (e.g., $n = 10$ or 20) the approximation was unsatisfactory. As this author points out, an important implication of these findings is that

> "The more meticulous one is in the analysis (i.e., the more terms one uses in the reliability equation), the lower the estimated reliability will be. It is apparent, then, that if different analysts choose a different total number of dimensions as being relevant to tasks, they will arrive at different estimates of the reliability of a system made up of these tasks." (p. 24)

2.2 Five Network-Based Techniques

The interest of human factors specialists in the formulation and use of network techniques for prediction of human performance derives from a number of sources. One is the observation that such techniques have proved useful for estimating the performance capabilities of systems composed of functionally interrelated items of hardware. A second is the realization that in most complex systems, a large number of functions are performed by the human and that the human can (and possibly ought to be) considered in the same terms as the hardware and software elements with which he interacts. A third source is the ease with which a network approach, combined with suitable mathematical operators, can accommodate a variety of data input types, from empirically-derived estimates of central tendency (mean or median performance) to theoretically-assumed distributions. A fourth and very important source, in the context of this report, is the hypothesis that a suitably designed network might serve as an important system-level mechanism for the aggregation of isolated and independent task and sub-task models currently available in the psychology, human factors, control theory, and other human performance literatures.

The goal of almost all network techniques is to get from a statement of the functional relations among individual task elements to an estimate of the performance of some defined aggregate of those elements. The performance parameters of interest are typically one or more of the following: (1) the time required to complete the task aggregate; (2) the probability that the task aggregate will be completed in a given time; (3) the probability that the aggregate will be completed successfully-

that is, at or above some level defined as acceptable. Because the techniques employ concepts similar to those of standard reliability theory, one can, if interested, utilize other performance measures such as mean-time-to-failure (MTTF), mean-time-between-failures (MTBF), etc.

The various techniques, though frequently quite different when viewed with respect to operating assumptions, degree of detail, etc., share common modelling requirements. All require a thorough description (frequently rendered in diagrammatic form) of actual or intended tasks and of the interrelationships among those tasks. Such procedural flows or operational sequence diagrams (OSDs) may sometimes be supplemented with other analytic devices such as the fault tree, which provide descriptions of the alternative ways a task or task-aggregate could fail to be accomplished in the desired time or at the desired level of reliability.

A second requirement is for "data," either in the form of measures of central tendency (the mean or median) or in the form of distributions, which can be taken as measures of performance in each of the tasks included in the aggregate. In some applications, empirically-derived measures or distributions have been employed, whereas in others, rational estimates or distributions thought to represent reasonable approximations have been utilized where empirical data are lacking.

As with data banks, no systematic taxonomy has been developed for defining or segmenting task sequences. What may be a task process with associated "data" at one level may be a small network of processes at another level, each with its own data which, in the aggregate, represent the task.

A third common requirement is for a set of rules and procedures for combining estimates of central tendency or distributions attributed to tasks into overall estimates or distributions representing the task aggregate. Frequently, the rules and procedures employed are those of standard reliability theory where, for example, the estimated reliabilities of separate components are multiplied together to yield an estimate of overall reliability, and the time required for each component to accomplish its function is added to that required by each other component in order to estimate total time to completion.

At least two problems have arisen in connection with this approach to human operator modelling. One results from a lack of empirical data relating to expected performance. In some cases the data simply do not exist. In other cases, the validity and/or reliability of available data are suspect. In either case, assumptions concerning expected values or likely distributions must be made, and these add an additional source of uncertainty to the modelling effort.

A second problem concerns the rules and procedures for combining point estimates or distributions. Where the aggregate contains some tasks that are highly cognitive in content - and particularly where a number of largely implicit tasks are conducted in parallel, either by the same or by different operators - the sum-of-times, product-of-reliabilities rules of conventional reliability theory have been found to be suspect. In addition, there are some as-yet-unanswered questions relating to the appropriate shapes of composite distributions made up of differently-shaped component distributions. These questions are, of course, of less concern in those approaches that utilize a Monte Carlo approach to generate the composite distribution, and in any case, are addressable through empirical research.

Another general observation one might make with respect to the network approach is that, while it may be suited to the modelling of discrete activities, it historically has not been suitable for modelling of continuous activities. Two points must be made in this regard, however: (1) Most networks were not developed with continuous tasks in mind; thus, failures of generalization are not appropriate criticism of the approach *per se*. (2) At least one network-based approach - SAINT - has been successfully extended to include certain families of continuous variables; thus, there seems to be no inherent limitation to continuous task modelling via network formulations.

We take the position in this report that networks are not, in and of themselves, models of human performance, though they may clearly be representations of system structures in which that performance is imbedded. As a result, we have included no abstracts of these techniques in the Appendix of this report. It must be recognized that this omission is not intended as a negative comment on their perceived value or utility in the assessment of system Performance, which, in our judgment, is quite high. The omission simply reflects the feeling that the various network techniques can be more usefully conceived as formal sets of rules and procedures for defining system structure and that these structures then provide the mechanism for aggregating data available from data banks or incorporating functional performance models.

In the remainder of this section we discuss four network techniques that have been applied successfully to the evaluation of human performance in man/machine systems. To this set we have added one other technique -- PERT -- which, while not a structure for evaluating human performance *per se*, provides the framework from which several current techniques are derived, and which serves as a convenient reference for discussion.

The set of four human performance network techniques selected illustrates quite well the variety of approaches currently available for prediction of performance. Further, the chosen techniques demonstrate what we consider to be the strengths and weaknesses of network methods. It is important to realize that we have not, by any means, exhausted the useful techniques in this class. For reviews of other techniques of the types illustrated here, the reader is encouraged to read Meister's (1971) comparative analysis of reliability models.

2.2.1 PERT

The application of PERT follows essentially the formula laid down above. *One begins by constructing a network consisting of* all major activities intervening between the start and the completion of a task. (Miller, 1963) Particular pains are taken to portray accurately all sequential dependencies implied in the task and to include activities that can or are intended to be conducted in parallel.

Following completion of the network, three estimates are made concerning each activity: (1) the earliest possible and (2) latest possible or allowable times at which it could be completed; and (3) the most likely time at which the activity would be completed. It is then assumed that these estimates represent the end points and mode, respectively, of a Beta probability density function with mean, t_e, equal to

$$t_e = \frac{t_o + t_p + 4t_f}{6}$$

where t_o = the earliest time

t_p = the latest time

t_f = the mode

and a variance, $\sigma^2_{t_e}$, equal to

$$\sigma^2_{t_e} = \frac{(t_p - t_o)^2}{36}$$

With expected times and variances thus derived for each activity, one then estimates the total task time, T_e, as the sum of the individual expected activity times,

$$T_e = \sum_{i=1}^{n} t_{e_i}$$

and estimates the variance in that time, $\sigma^2(T_e)$, as the sum of the individual activity variances,

$$\sigma^2(T_e) = \sum_{i=1}^{n} \sigma^2_{t_{e_i}}$$

One can also estimate the probability that the task will be completed as of a specified time - frequently, a desired project end-date. To accomplish this, one assumes that the composite of activity completion times has a normal distribution and then computes the ratio of the area under that distribution to the left of the specified time to the area under the total distribution. In units of standard deviation (σ), this becomes

$$\frac{T_s - T_e}{\sigma}$$

where T_s = the specified time and T_e is defined as above.

There are, of course, many other facets to a PERT analysis. Ordinarily one is interested in such things as the critical (that is, the longest) path through the network, the distributions of positive and negative "slack" among activities, etc., for purposes of controlling a total effort. What is of major interest here, however, are the strengths and weaknesses of that portion of the model sketched above.

This technique has two primary strengths. First, there can be little doubt that the development of a network in detail sufficient for the application of PERT forces one to think critically about the relationships among activities. In specifying activities at a fine level, the analyst can often determine in some detail what he does not know and what information he must acquire to assess the task.

Secondly, the technique provides quantitative information useful for the assessment of project completion time and provides a framework for assessing performance reliability. With this information and framework, one can hope to characterize the responsiveness of an intended mission and to consider alternative configurations.

The weaknesses we see have mainly to do with the means by which the temporal data characterizing an activity are acquired, the distributions assumed, and the mathematical simplifications employed. As indicated, the analysis requires estimates of earliest, latest, and modal times. These frequently represent either the opinions of individuals familiar with activities similar to those under study or the outputs of part-task studies or laboratory experiments.

There seems every reason to expect that the pattern of interaction that gives rise to the expert's estimates or the experimental results will be different from those that will operate in the task being assessed, unless one is so familiar with both contexts as to be able to ensure the validity of the estimates being used for analysis.

In addition, much concern has been shown in the literature over the assumption that the time estimations employed in PERT are fitted into a Beta distribution with a variance equal to one-sixth of the range between the earliest and latest times. Three possible sources of error arising from this assumption have been studied empirically by MacCrimmon and Ryavec (1964), who found that errors of 30 percent and 15 percent can be introduced into the calculation of the expected time and standard deviation as a result of failures to meet this assumption. They also found that approximately the same sorts of error rates would have been observed with a much simpler triangular distribution. Thus, although the Beta-distribution has strong intuitive appeal, it may be open to question on empirical grounds.

The assumption concerning normality has more in its favor, since it is based on the Central-Limit Theorem. The critical question in any particular application, of course, is whether or not there is a sufficient number of activities in a sequence to justify the assumption that the distribution of sample means is approximately normal. Meister (1964) mentions in this regard an unpublished study by Rook, who found that a distribution of failure probabilities may be considered normal if the number of parameters exceeds 15.

Finally, there is the question of whether or not the mean and variance of the total task time are equal to the sums of the individual means and of the individual variances, respectively. Research on this point has suggested that the answer depends critically on the relative lengths of paths in the network and the number of activities shared in common by different paths. This is of great concern, it seems to us, for it suggests that even when one can be confident of his temporal data and of its distribution, the very nature of the task structure may preclude a precise analysis.

It should be apparent that the MacCrimmon and Ryavec analysis is related only to temporal aspects of the models assumptions and not at all to the concept of reliability therein. Since, however, the accuracy of the reliability assessment is based entirely on estimates of total allowable task time and on the derivation of expected time, errors in computed reliability can be expected to be monotonically related to errors in total expected project time.

2.2.2 The Siegel and Wolf Technique

The Siegel and Wolf Model (SWM), which appeared in 1961, has been of continuing interest to human performance analysts and recently has been extended to the study of survivability/vulnerability in nuclear environments (Chubb, 1971). It embodies some of the network concepts of PERT and, like that earlier formulation, depends significantly on relationships between required task times and available task times for measures of completion probability and stress. Unlike PERT, however, it is a human performance prediction technique and, even more significantly, a technique employing Monte Carlo simulation. It is also built upon conceptual structures and empirical observations drawn from psychology and human factors and, as such, can be considered a model in its own right.

An indication of the analytic interests represented by SWM is provided by the following quote:

> "It is the purpose of this model to give the equipment designers quantitative answers to questions such as the following while the equipment is in the early design stage:
>
> (1) Can an average operator be expected to complete successfully all actions required in the performance of a task within a time limit, T, for a given operator procedure and a given machine design?
>
> (2) How does success probability change for slower or faster operators and longer or shorter periods of allotted time?
>
> (3) How great a stress is placed on the operator during his performance and in which portions of the task is he overloaded or underloaded?
>
> (4) What is the frequency distribution of operator failures as a function of various stress tolerances and operator speeds?" (Siegel & Wolf, 1961).

To apply the model, one first performs an analysis and identifies which tasks are essential to completion of the mission and which are unessential. For each of these, the following data are specified, using "the best available sources":

(1) The average time required by the operator to perform each task.

(2) The average standard deviation about the average time.

(3) The average probability of successfully performing each task.

(4) An estimate of the extent to which successful performance of each task is required for completion of the total mission.

(5) The waiting time (if any). This represents the time elapsed between the start of the mission and the start of each task, during which no action can be taken by the operator.

(6) The next task to be performed given failure to accomplish a current task.

(7) The next task to be performed given successful accomplishment of a current task.

Following the determination of these values, one then computes urgency and stress conditions, which are defined with respect to the available time and sequential constraints of the network. (See Abstract Nos. 2 and 3.) A situation is "non-urgent" when there is sufficient time to complete all remaining subtasks, assuming average speed and error-free performance. A situation is "urgent" when only enough time remains to complete the required tasks, and it is "highly urgent" when insufficient time remains to complete these tasks. When the situation is

"non-urgent" or "urgent," one assigns a stress level with a value of unity. When it is "*highly urgent*" the level is defined as the ratio of the sum of the average time required to complete remaining required tasks to the total available time remaining in the mission*

$$s_i = \frac{\bar{T}_i^E}{T - T_i^U} \geq 1$$

where s_i = the stress level

$\bar{T}_i^E$ = the average time required to complete all remaining required tasks in the mission

T = the total time available for completing all tasks in the mission

T_i^U = the total time required to complete all tasks through $(i - 1)$

An important part of SWM is its incorporation of the psychological concept of a "stress threshold." As the authors note, "Theory suggests that emotion or stress up to a certain point acts as an organizing agent on behavior; beyond this point stress acts as a disorganizing agent" (1961; p. 20). The model simulates this effect by comparing the value of s_i against a value M, input by the user prior to each simulation run, as explained below.

*For purposes of simplifying this presentation, where possible, we shall follow the Siegel & Wolf (1961) notation, rather than that contained in later publications.

With the stress levels computed, one then computes the time required for completion of the subtask, based on the assumption of a normal distribution with mean and standard deviation equal to those specified earlier. A specific value for a given "exercise" of the model is chosen via a Monte Carlo technique that samples from the distribution of values assigned to each subtask. The computational procedure utilized generates a sample value in which the average completion times and average deviations:

"(1) are used unchanged when stress equals unity;

(2) are decreased with increasing stress until stress assumes the threshold value;

(3) are used unchanged when stress equals the threshold value;

(4) are increased linearly with increasing stress beyond M until, when stress equals M+1, the contributions of $\bar{t}_i$ (average completion time) and $\bar{\sigma}_i$ (average standard deviation) remain constant at $2\bar{t}_i$ and $3\bar{\sigma}_i$, respectively." (Siegel & Wolf, 1961). Note: $\bar{t}_i$ and $\bar{\sigma}_i$ are the average time to complete task i and the average standard deviation of the average time, respectively.

Finally, one determines whether or not the task can be considered to have been completed successfully. Critical to the computation here are the nominal probability of success specified earlier, the stress level, and the stress threshold. The computational rule is as follows:

$$p_i = \begin{cases} \bar{p}_i + \dfrac{(1-\bar{p}_i)(s_i-1)}{M-1} & \text{for } s_i < M \quad \text{(see footnote)} \\ \bar{p}_i(s_i+1-M)+(M-s_i) & \text{for } M \leq s_i \leq M+1 \\ 2\bar{p}_i - 1 & \text{for } s_i > M+1 \end{cases}$$

where p_i = probability of successful performance of task i

$\bar{p}_i$ = average probability of successful performance of task i

s_i = stress level of task i

M = stress threshold

As Siegel and Wolf note, the probability of success thus defined increases in linear fashion with stress, reaching unity at the stress threshold. At this value, it assumes the average (i.e., input) value, $\bar{p}_i$, and then decreases linearly until it reaches a constant value.

Since the outcome must be considered to be a binary event, a decision as to whether the task was, in fact, completed successfully requires the comparison of p_i with a pseudo-randomly generated number, R. The decision rule specifies successful performance if $p_i > R$; otherwise, performance is considered unsuccessful.

*An error in the specification of the multiplier is contained in the 1969 version of this argument. The correct form, analogous to that given here, is

$$\bar{p}_{ij} + \frac{(1-\bar{p}_{ij})(s_{ij}-1)}{M_j-1}$$

In actual use, the computations alluded to above are performed by computer, which also keeps track of the total time remaining for performance of the mission. When that time has been exhausted or when the task has been completed successfully, the simulation is terminated. At the conclusion of a run, data pertinent to task completion times and reliabilities are recovered for analysis.

An important aspect of SWM is its suitability for use in simulations of the performance of multi-person systems. In this role, the technique employs a number of self-contained models to determine expected task time as a function of group performance proficiency, overtime load, morale, number of persons, and nominal (average) time.

Some feeling for the amount and specificity of data required as inputs to the Siegel and Wolf group simulation is provided by the authors' listing of "mission data" that must accompany the definition of each action unit (task) in a total crew assignment:

(1) The average time h_{ud} in tenths of hours, required to complete action unit u on day d.
(2) An indication of whether or not h_{ud} represents a fixed time period, a variable time period, or a time-limited period. A time-limited period is one that must end by a specified time of day regardless of the starting time (variable = 0, fixed = 1, limited = 2).
(3) An indication of whether or not the action unit performance is to be repeated or improved in the event that performance efficiency does not come up to expectation (no repetition = 0, repetition or touch up = 1).

(4) An indication of whether or not communication between stations is required on this action unit (yes = 1, no = 0).

(5) The carry-over code that specifies the importance of the action unit:

essential = 1
postpone to avoid overtime = 2
postpone if crew morale is below the threshold = 3
ignore to avoid overtime = 4

(6) The orientation values of the action unit:

benefit to individual, S_{ud}
benefit to crew, C_{ud}
benefit to mission, M_{ud}

(7) The type of action unit--an indicator of the preference in selecting group members:

normal, prefer primary specialty = 1
training, prefer alternate specialty = 2
difficult, insist on primary specialty = 3

(8) A prior action unit number q_{ud} that must be completed before action unit u can be initiated.

(9) Indicators e_{ud} to specify for each equipment system whether or not it is required to accomplish the action unit (yes = 1, no = 0 for each equipment, e). (Siegel & Wolf, 1969, pp. 71-72).

In addition to these, the model makes provisions, where relevant, for use of a variety of input data relating to equipment used (failure rates, repair times, etc.), to personnel parameters (training and operational experience, proficiency, etc.), to crew cohesiveness and morale, etc.

The ability of the group simulation model to predict empirically and independently verifiable outside criterion data has been tested in the context of a realistic nuclear missile submarine training mission (Siegel and Wolf 1969). This check on the validity of the model had a number of specific purposes: (1) to compare the model predictions of required crew composition with results obtained from field experience; (2) to compare model predictions of average crew proficiency, supervisor satisfaction, and number of unessential work units remaining unperformed with estimates of those parameters by operating personnel; and (3) to assess the internal consistency of results relating crew efficiency, productivity, cohesiveness, and morale as a function of crew size. The predictions made by the model appeared to accord well with the expectations and predictions of FBM personnel, indicating, from the authors' point of view, a generally valid model development.

In addition to the benefits derived from analyzing operating procedures in sufficient detail to develop the network representation, SWM has a number of significant virtues. One is a mechanism for modifying performance in accord with stress. A second is the Monte Carlo component of the technique. Rather than predicting a judgment of the performance of the system on the sum of single subtask expected values, as in PERT, SWM expressly considers that operators will differ in their performance of the same subtask and that the same operator will exhibit differences in successive repetitions. This capacity to encompass both within-operator and between-operator variability is, in our judgment, an important desideratum in performance modelling. A third is the ability to address multi-person systems problems.

Weaknesses that we see are similar in some respects to those encountered in PERT. There is, first of all, the question of whether valid structural representations and data can be assembled for a system in which possible interactions may be only dimly understood. Secondly, there is a question concerning the soundness of the assumption that the means and standard deviations specified are, in fact, the means and standard deviations of normal distributions. Though such an assumption might be justified, as in PERT, in terms of the Central Limit Theorem, results of a very recent study by Mills and Hatfield (1974) suggest that the three parameter Weibull (or weighted Weibull) distribution may provide a better fit.

Thirdly, we have some concern over the ability of SWM — or, indeed, any network technique — to yield accurate predictions of performance in situations where observable task density is low (as, for example, in tasks requiring a great deal of monitoring and signal detection but only occasional system input) or where the performance of several operators functioning in parallel must be assessed. The first of these situations is of concern not only because the network itself is difficult to construct when activities are largely implicit, but also because sufficient characterization of those activities for purposes of assembling parametric input data may not be achievable. Our concern in the latter situation is that the reliability of the estimate may, as in PERT, depend on the network configuration and the degree of interconnectedness of the tasks.

2.2.3 SAINT

Unlike PERT, SAINT (Systems Analysis of Integrated Networks of Tasks) is a network technique that embodies a variety of types of process branching, a set of alternative distributions for use in modelling individual task elements, and a Monte Carlo procedure for sampling from the distributions specified. Designed to capture the best features of the Siegel and Wolf approach and the networking and symbol manipulation capabilities of GERT (Graphical Evaluation and Review Technique; see Pritsker and Happ, 1967), SAINT contains a Fortran-based user language and a formal symbolism within which to represent and code critical parameters at each network node.

SAINT has been employed in a wider range of system modelling contexts than any other network approach identified in this review. The following are illustrative:

1) Choice reaction time (Hann and Kuperman, 1975)
2) Remotely piloted vehicle/drone (RPV) control facility (Wortman, Duket, and Seifert, 1975)
3) Digital Avionics Information System (DAIS) design (Kuperman and Seifert, 1975)
4) Hot strip mill process control (Maltus and Buck, 1975)
5) Airborne warning and control systems (Mills, 1976)

In this section we shall attempt to highlight those characteristics of the technique that appear to play major roles in its utility and generality. Readers interested in specific applications and in the design philosophy and internal organization of SAINT are urged to read the above references and the system documents provided by Pritsker et al (1974) and Wortman, et al (1974), respectively.

2.2.3.1 Task Concepts

As in PERT and SWM, tasks are related to each other in SAINT by precedence relationships. These precedence relationships specify the flow of operations through the network and indicate which tasks can be started as a result of the completion of given prior tasks. Unlike PERT, the completion of particular tasks in a SAINT network can modify precedence relations among later tasks, thus altering the flow through the network.

A given task has associated with it input, task, and output parameters that specify the number and nature of predecessor tasks that must be completed before the task in question can be begun, characteristics of the task and statistical information to be collected and the branching (to other tasks) to be performed at its conclusion, respectively. A brief summary of important parameters, in addition to those that specify flow through the network, is presented below for the purpose of highlighting the modelling flexibility that is achievable through the SAINT technology:

2.2.3.2 Task Duration

The time to perform a task is specified in terms of a desired sampling distribution and a parameter set identified by the user. The following distributions are available:

1) Constant
2) Normal
3) Uniform (Rectangular)
4) Erlang
5) Lognormal
6) Poisson
7) Beta
8) Gamma
9) Beta fitted to three values
10) Scaled constant
11) Triangular

If none of the above functions are appropriate to a given task, but an alternative can be specified, the SAINT language provides a means by which the user can write the alternative function as a subroutine and employ it in the simulation.

2.2.3.3 Task Essentiality

SAINT provides a mechanism whereby tasks may be graded in terms of their essentiality to the total system mission. When, during simulation of the mission, remaining time grows critically short, less important tasks may be skipped in order to assure the completion of essential tasks.

2.2.3.4 Task Type

Six possible task types can be defined in SAINT:

1) Single Operator Task
2) Joint Operator Task
3) One of several operators (first to become available completes job)
4) Hardware Task (no human operator)
5) Cyclic Task
6) Gap Filler Task (completed only if time available)

2.2.3.5 Task Output

After a task has been completed, a decision is made as to which of n tasks should next be started. Since the set of possible next tasks is defined by the network, this decision results in the selection of m from among the n possible branches connecting the completed task with remaining tasks. Five decision rules are implemented in SAINT:

1) "Deterministic," in which all n branches are selected.
2) "Probabilistic," in which one branch is selected on the basis of a random number drawn from a uniform distribution.
3) "Conditional, 'take first'," in which the first branch staisfying a specified condition is selected. Possible conditions are (a) associated task complete; (b) accrued simulation time equal to or less than a specified time; (c) associated task not completed; (d) accrued simulation time greater than a specified time.
4) "Conditional, 'take all'," similar to the preceding rule except that all m branches satisfying stated condition are taken.
5) "Modified probabilistic," similar to "probabilistic," except that branch probabilities are qualified by the number of previous completions of the task from which the branches emanate.

2.2.3.6 Statistics Collected on Tasks

Five types of statistics relative to temporal aspects of performance can be collected and tabulated in SAINT.

1) Time of first completion of a given task
2) Time of all completions of a given task
3) Time between (successive) completions of a given task
4) Time required to complete a given task
5) Time elapsed from completion of the first predecessor task to the start of a given task.

In addition to those parameters that characterize task structures in SAINT, there is a set that characterizes the operator(s) and that serves "to make a mission operator-specific" (Pritsker, et al, 1974; p. 17). The essential members of this set are summarized briefly below:

2.2.3.7 Operator Speed

Values assigned to this parameter reflect differences in task performance time caused by differences in proficiency and training among individual operators and groups of operators.

2.2.3.8 Operator Accuracy

Values assigned to this parameter reflect differences in performance accuracy due to differences in proficiency and training among individual operators and groups of operators.

2.2.3.9 Operator Stress Due to Workload

As in SWM, the stress imposed on an operator is defined as the ratio between the time required to complete remaining tasks and the time available to complete them. In addition, the basic SWM principles associated with performance accuracy and time as a function of stress level are contained in SAINT.

2.2.3.10 Goal Gradient

The commonly observed increase in the nominal accuracy of performance as one nears completion of a task can be simulated. To approximate this effect, the SAINT user prescribes, for each modelled operator, a value reflecting the percentage of mission completion required before onset of the gradient. The gradient itself is produced by adding to the nominal probability of success a value whose size is a direct function of the extent of prescribed task completion.

2.2.3.11 Comment

As a simulation utility that employs a bottom-up approach to performance prediction, SAINT is probably without peer at this time. It incorporates what we consider to be the most satisfactory concepts with respect to task and operator parameters identified in SWM, and employs a high level language that is easily learned and manipulated by the user. Further, the very flexible branching structure and the capability for changing the sequence of subsequent tasks offer what is perhaps a unique opportunity for the simulation of system missions with broad dynamic range.

Like other network approaches, however, the validity of SAINT's predictions is only as good as the completeness and validity of the set of individual performance models underlying them. Where in a real mission, for example, stress is only partly a matter of required tempo, SAINT prediction's concerning task branching and performance accuracy may be misleading. The point here is an obvious one and has been made by us and by others in connection with other bottom-up approaches. The matter is, nonetheless, of continuing concern.

2.2.4 Lockheed-Georgia

Because it is oriented toward establishing the safety and reliability of systems, the Lockheed-Georgia technique (L-G) is concerned explicitly and almost entirely with the identification and assessment of failure modes in complex man-machine systems. It represents an analytic approach to performance prediction as opposed to the synthetic approach employed by the three techniques reviewed above. In the study of alternative configurations of men and equipment in All-Weather Landing Systems (AWLS), from which the material for the current discussion was drawn, the technique employs point estimates of reliability selected from data banks, the human factors literature and, in one instance, a model of human scanning and detection performance (Adkins et.al., 1967). The technique employs the concept of fault-trees and utilizes the standard conventions of reliability theory for aggregation of reliability data contained in those analytic structures.

The Lockheed-Georgia AWLS technique is intended to answer the following questions:

"1. Do the demands that are placed on the pilot in terms of oral, visual, and physical requirements at the critical phases of landing and the multiple task situations impair safety or have disorientative effects on the flight crew?

2. What is the adequacy of the visual command and position information as well as of malfunction clues?

3. What is the distribution of work load among crew members?

4. Is the visual scanning pattern optimum such that maximum error-free assimilation of data can be achieved? Is the location of instruments and controls optimum?
5. Are redundancy and/or monitoring provisions presented in a manner that accomplishes the desired objective? Are these provisions effective, and usable? As an example, the copilot flight direction information may serve as backup for that of the pilot, but is there any effective redundancy if the copilot is looking out the window in an attempt to establish visual contact with the ground while the pilot is totally concentrating on his own instrument system?
6. What is the probability of inadvertent actuation or response for a given task? Also, what is the susceptability to misinterpretation or misreading of visual displays?
7. What is the adequacy of the feedback as a result of pilot action during the critical phases of flight?
8. Does the sequence of tasks being performed during the landing operation yield maximum efficiency and minimum error probability on the part of the crew?"

As in almost all techniques of the network type, the first step in the L-G approach is a comprehensive analysis of the tasks and subtasks that must be accomplished, the time required to complete each, and an assessment of the criticality of successful completion. Since in the AWLS project, the displays and controls employed in the tasks/subtasks were of direct concern, this analysis included a study of "informational input" and "control output/feedback" to the pilot and crew.

From this task analysis and a parallel analysis of system hardware and control functions, fault trees were constructed that depicted the ways in which the mission (in this case, a successful landing or successful go-around) could fail, and the etiology, in terms of malfunctions of equipment and incorrect responses on the parts of operators, associated with each of these failure modes. Reliability data were then drawn from a variety of data banks (e.g., AIR Data Store, Titan II, Aerojet General) and posted to appropriate branches of the trees.

As noted above, the degree of success with which various displays and controls could be used by operators was considered to be critical to mission outcome. Further, it was considered that degree of success was related monotonically to the location of the required display or control on the instrument panel. Given this assumption and the obvious impossibility of locating all instruments at the most desirable location (panel center), an empirical model based on detection and response probabilities as a function of off-axis scanning provided inputs to branches associated with failures in input or feedback of information.

The nominal probability of occurrence of a given failure mode identified in the tree was then assessed as unity minus the product of success probabilities posted to branches leading to that mode. In the AWLS application, however, it was important to consider that occurrences of given failures might only be critical to success if they occurred at certain times in the landing sequence. In such instances, rate of failure production (as related to duty time, mean-time between failure, etc.) became important. Where relevant, then, probability of success is defined by the standard reliability equation.

$$P_r = e^{-t/m} \qquad \text{where } m = 1/f$$

and P_r = probability of success
t = duty time
m = MTBF
f = failure rate

There are many other facets to the L-G approach that make it useful in the assessment of complex aggregations of procedures and hardware. One is ASSET (Advanced System Safety Engineering Techniques), which contains a large set of programs for computing the reliabilities of different types of series, parallel, and redundant man/machine networks and of deriving from their nominal reliabilities an estimate of the associated "safety" level, based on load factors, approach patterns, ground control accuracy, etc. A second is the transformation of historical accident data into "hazard" information that can be used in assessment of the criticality of various failure modes.

2.2.5 THERP

Another tree-structure approach to performance prediction is the Technique for Human Error Prediction (THERP) developed by Swain at Sandia. This technique has been discussed in a variety of different contexts (see, for example, Swain and Guttman, 1975; Swain, 1964; 1973; 1974) and has been reviewed at length by Meister (1971).* In view of this exposure, we shall discuss only what we consider to be the highlights of the approach.

It is interesting to note that the developers of THERP do not consider their approach to be a model in the classic sense:

> "The term 'model' has several meanings, and is used loosely in the behavioral sciences. For our purposes in describing THERP (Technique for Human Error Rate Prediction) we will use the narrow meaning of 'model,' defined as 'a set of relations and operating principles.' Although this may slight some who are interested in theory development, THERP was developed as an applied technique, for use as a fast and relatively simple method of providing data and recommendations to systems designers and analysts who require a quantitative estimate of the effects of human errors on system performance.
>
> "THERP is not a model in the usual sense of a hypothetical analogy, although it is regarded as a modest form of mathematical (Boolean) modeling, representing behavior variability by simple equations dealing with equipment parameters, human redundancy, training, stress, etc." (Swain & Guttman, 1975, p. 116.)

---

*An up-to-date bibliography of Sandia reports on THERP is contained in the Sandia Corporation publication, "Description of Human Factors by Sandia Laboratories," Sandia Corp., Albuquerque, New Mexico, January, 1976.

As noted by Meister (1964), THERP is an iterative procedure consisting of the following five steps (which may or may not be accomplished and re-accomplished in the same order, until system failures resulting from human error are at an acceptable level):

"(1) Define the system or subsystem failure which is to be evaluated.
(2) Identify and list all the human operations performed and their relationships to system tasks and functions.
(3) Predict error rates for each human operation or group of operations pertinent to the evaluation.
(4) Determine the effect of human errors on the system or subsystem failure rate as a consequence of the estimated effects of the recommended changes." (p.630)

As in all other network techniques, the first two of these steps aim to formalize the tasks under study and to yield a specification of their interrelationships. The outcome of these prior activities is a diagrammatic task analysis in the form of a tree structure that represents subtask elements as nodes and probabilities of successful and unsuccessful accomplishment as branches. The data for these later estimates came originally from the AIR Data Store but more recently represent estimates derived from empirical research conducted at Sandia Corp. (Swain, 1974.)

When the tree structure has been completed and probabilities have been posted to its branches, the branch estimates are aggregated in order to derive a measure of system or subsystem performance. The estimate of interest here from a systems point of view is the probability that if a failure occurs (either in the hardware or human element or in both jointly), that failure will result in degradation of system performance.

2.2.5.1 Prediction and Aggregation of Error Rates

The assignment of branch probabilities and their aggregation are critical activities in the exploitation of THERP. As we have said, information from data banks represents an important resource in the first of these tasks, but the retrieval, selection and application of this information is indirect and highly judgmental in character. Swain (1964) points out that the "predictions" of error rate posted to branches in the probability trees are ordinarily made by a team of project engineers, operations researchers, safety engineers, and human factors specialists who, in the aggregate, are intimately acquainted with the detailed functioning of the system being modelled, the equipment being used, environmental factors obtaining, etc., and with the interactions that may occur between system events. Furthermore, a variety of different values, from one or more point estimates of average rate, to "fiducial" points reflecting, for example, estimates two or more standard deviations from the mean, are predicted, depending on the requirements of the problem.

After error rates are determined, estimates are then made, again by experts, that errors of the specific type or class identified could result in system failure. The probability of a failure is then modelled as the product of the probability (F_i) that each error identified, if it occurs, will result in the failure and the probability (P_i) that a system operation permitting the error in question to be made will occur.*

*Swain points out that the definition of P_i may vary, depending upon the application: "...P_i may result from P_1 or P_2 where P_1 and P_2 are human errors. Thus $P_i = P_1 + P_2 - P_1 P_2$. Or, P_i may be the joint probability of P_1 and P_2, where P_1 is an error and P_2 is an equipment defect or some other factor which, when it occurs, sets up a potential failure condition only if a human error (P_1) also is made. Thus $P_i = P_1 P_2$. (P.632).

The total system failure rate resulting from errors in human performance is given by the expression

$$Q_t = 1 - [\prod_{k=1}^{n} (1-Q_k)]$$

where Q_t = the probability that one or more failure conditions will result from errors in at least one of n classes of errors, and Q_k is the probability of one or more failure conditions that exist as a result of Class k errors occurring in n operations.

2.2.5.2 Assumptions Concerning Human Performance

THERP makes a number of important assumptions concerning human behavior and performance. *Among them are the following:*

1) The standard deviations of the distributions of error rates for two tasks that are similar are proportional to the means of the distributions for those tasks.

2) Where two operators work together, the reliability, R, due to redundancy is given as

$$R = \frac{1-(1-R_1)\ (T_1) + R_1\ (T_2)}{T_1 + T_2}$$

where R_1 is the reliability of a single operator, T_1 is the percent of time the second operator can observe the first and T_2 is the percent of time remaining to complete the task (Meister, 1964). The probability of one person's detecting another's error when error frequency is low is equal to 0.85.

3) Denoting the probability of error in an important time - critical task on which errors occur infrequently as p, the probability of making an error on n repetitions of the task is equal to 2p(n-1), up to a limiting value of 1.0.

4) For conditions of little time stress, the probability of error on the first repetition can be approximated as the square of the probability of error on the first trial.

2.2.5.3 Comment

Used by human factors engineers intimately familiar with the design and operation of the system whose performance they are attempting to predict, THERP and the L-G approach appear to be successful techniques. The concentration of analytic activity prior to the actual construction of the network and aggregation of assigned probabilities serves the extremely important function of aiding the identification of system missions that could easily be compromised by unreliable performance of men, equipment, or their interaction.

It is important to note that there are a number of pitfalls associated with the fault tree approach, particularly when it is employed in the context of an incompletely specified system. Fussell (1974) has accumulated these into three basic categories:

1) oversight and omission of potential failure modes and their outcomes,
2) application of poor or inapplicable failure data or of incorrect assumptions to highly complex systems, and
3) failure to account properly for mutually exclusive events that occur in the same tree.

To the extent that these pitfalls can be avoided through the involvement of knowledgeable personnel and with the computer aids to fault tree analysis that are beginning to appear (see, for example, Vesely and Narum, 1970), this technique represents a valuable tool for performance prediction.

2.2.6 Summary of Network Techniques

In this section we have briefly reviewed five performance prediction techniques that employ network methods. In the aggregate, these techniques illustrate the following:

1) The use in analytic and simulation approaches, respectively, of point estimates and of distributions of task completion time and reliability.

2) The use of data bank information and/or information generated from individual task performance models.

3) The use of three different concepts of reliability: one based on the ratio of total required task time to total available time; a second based on empirically determined probabilities of successful (or unsuccessful) performance with no explicit time constraint; and a third based on estimates of mean-time-between failure and required duty time.

4) Two essentially different approaches to the assessment of reliability. One of these is synthetic in nature, requiring representation in network form of all tasks and activities that must be accomplished in order to complete a mission. The other is analytic, requiring representation of those tasks that, if accomplished incorrectly, will result in an unsuccessful mission.

5) The use of qualitative contextual information to sharpen time and reliability estimates made on the basis of historical data from the experimental literature or field experience.

This set also illustrates the difficulty of maintaining a distinction between "models" of the type that form the major substance of this report and what we have chosen to call "network techniques." At least one of these — SWM — is a legitimate member of both categories.

In the discussions above, we have made a number of comments regarding characteristics of the network approach. It is appropriate here to summarize these points:

1) There seems little doubt that consideration and evaluation of a system process in enough detail so that procedure networks and/or fault trees can be constructed is of value in highlighting aspects of performance that are *critical to mission success*. The value of this activity would, in our judgment, be high even if no subsequent attempt were made to employ the networks in a predictive fashion through the aggregation of branch reliabilities and/or completion times. From this point of view, the network approaches have the same general heuristic value that has been assigned to the use of decision trees in medical diagnosis (see, for example, Lusted, 1968) and flow charts in computer programming.

2) Except where directly relevant experience with similar tasks in similar contexts exists, distributions of within operator and between operator performance times and reliabilities must be assumed in the characterization of subtask components. There are few principles to guide a choice among possible distributions or to guide a set of choices in the event that different subtasks might be thought to have different distributions.

3) In those models where point estimates are combined over a set of subtasks, there are few principles to guide the selection of a composite distribution, except, possibly, the Central Limit Theorem. (Even here, there may be some doubt concerning the number of individual activities required for a valid assumption of normality.)

4) Even where one has considerable confidence in either point estimates of distributions of subtask completion time and/or probability of success, the nature of the interrelationships among subtasks in the network may preclude accurate prediction of total task parameters. The matter is of only minor concern where a single individual may perform more than one operation at given time (*e.g., as in a pursuit tracking task*) but is of significant concern where multiple operators perform in parallel. At the very least, one must be concerned about the combinatorial rules used to aggregate subtask data, and be prepared to find that the product and summation rules drawn from standard reliability theory do not afford adequate approximations.

5) There are no direct mechanisms within the models by which to assess such factors as within-subtask learning and time-sharing.

6) The approaches may not yield acceptable predictions when the observable subtask density is low and/or is composed of activities which are largely cognitive in content.

Two points must be made concerning these alleged difficulties. First, most can be addressed experimentally by comparing the actual performance of operators with the predictions of the models. An example of this approach is provided by the work of Mills & Hatfield (1974). Following this approach, one could examine the consequences for total task prediction of network complexity, network density, and subtask character (implicit vs. explicit performance) and could then derive appropriate combinatorial rules under different conditions of time stress and operator sophistication.

Secondly, one might find that certain simplifications (such as the assumption of normality or use of a product rule) do not introduce errors of sufficient magnitude to invalidate their use.* As Knowles, et al (1969) have noted,

> "Fortunately, the 'if-then' nature of system-level evaluations based on analytical models early in the design cycle suggests that a high degree of absolute precision in human performance data may not be a prime requirement. In most cases system configurations are compared with one another to determine the best of several potentially acceptable designs. In the rest of the cases the problem is to arrive at a ballpark estimate of system performance, or to allocate performance requirements so that potential trouble areas can be identified, or to set criteria for suboptimization of parts of the system." (p. 581).

*A case in point occurs in a study by Adkins, et al (1967) of the reliability requirements for all weather landing systems. These authors were able to show that almost no matter what gratuitous assumptions were made about the performance of the pilot and co-pilot, the reliability afforded by a particular configuration of auto-pilots made mandatory their choice as the prime control element.

2.3 Control Theoretic Models

By control-theoretic models we mean those models that have their analytic bases in control theory and statistical estimation and decision theory. Traditionally, these models have attempted to describe human performance in continuous control and monitoring tasks. The emphasis has been on models that describe the human operator as a system element in a feedback control loop. In this context, models exist for continuous control (manual control), for instrument monitoring, and for detection of and adaptation to abrupt changes in a system (such as would occur in a failure). In addition, approaches to incorporating stress effects in these models, especially those stresses associated with attentional demand (i.e., workload), have been developed.

Control- and decision-theoretic models could be useful in the analysis of various phases of command and control problems such as those associated with RPV control. For example, the problems of monitoring and controlling separation between RPVs, including those imposed by sharing attention in the control of several RPVs, may be amenable to analysis with such models. Of course, the most likely area of application in command and control would be in problems of continuous control, e.g., the "terminal" phase of the RPV mission.

Models of this type have been developed largely by systems engineers with a primary interest in predicting total man-machine system performance. This history has had a direct impact on the characteristics of the models that have emerged. Thus, the models are not intended to provide structural analogs of the human operator, though some do incorporate a degree of structural similarity. The models tend to describe human performance at the system level and in system terms. Concern is more with the operator's continuous

interaction with the system, as demanded by closed-loop analysis, than with his response to discrete events.

The system approach used in manual control models has implications for the level of detail at which man-machine interface problems are addressed. In the display area, for example, the principle concern has been with the information requirements of a task rather than with the human engineering aspects of display design. Thus, applications of the models have generally addressed the question of whether the information provided to the controller is adequate to achieve requisite performance with acceptable workload and, if not, what information is necessary. The effects of display layout on performance and workload are also treated (by the visual scanning models). However, display questions related to display readability such as color, brightness/contrast, linear vs. rotary indication, inside-out vs. outside-in, etc., are largely ignored in manual control analyses. Though there has been some recent work on control stick design using manual control models (e.g., Levison and Houck, 1975), detailed analysis and design of this aspect of the system interface has been missing.

Control-theoretic models generally require as an input a relatively detailed, analytic model of the "machine" to be controlled and/or monitored. This occurs because the ultimate form of the model or its parameters reflects human adaptation to the task to be performed and, furthermore, because these models are generally appropriate only for skilled operators who are attempting to act rationally to achieve well-defined objectives. Finally, we note that the models have been applied principally in the area of vehicle control. Only recently have the model structures and notions been developed that suggest the potential for application to problems in which human control

actions are infrequent and in which monitoring and decision-making are the operator's main activities.

2.3.1 Models for the Human Controller

Sophisticated mathematical tools have been employed in an attempt to develop analytic or computer models that will predict accurately the performance of human operators in continuous control tasks and a number of different models have resulted. These include, for example, sampled-data models (e.g., Bekey, 1962), finite-state models (e.g., Fogel and Moore, 1968), learning models (Preyss, 1968; see Abstract No. 4), and so on. Although many of these models have interesting and useful features, it is fair to say that, at present, the manual control field is dominated by models based on control theory developments, namely, quasi-linear describing function models, fixed-form parameter optimization models, and state-space optimal control models. This review will focus on these dominant models. However, because of their potential for incorporation in total mission models, simulation models for manual control will also be discussed (including simulation versions of the dominant models).

2.3.1.1 Quasilinear Models

The human controller is a complex control and information processing system. It is generally acknowledged that he is adaptive, time-varying, frequently non-linear, and that his behavior is stochastic in nature. Such systems are, of course, very difficult to analyze and characterize. However, there has been considerable success in applying quasilinear describing function theory to the problem of modeling human control behavior.

The development of quasilinear theory and its application to manual control was pioneered by Tustin (1944), (1947). This work and that of Russell (1951), Elkind (1956), McRuer and Krendel (1957), and McRuer, Graham, Krendel and Riesner (1965) resulted in a set of quasilinear models that are quite adept at predicting human control behavior in the simple, but important class of problems involving compensatory tracking with a single display and a single manipulator. The work is well summarized in the excellent report of McRuer, et al. (1965) and in several surveys (e.g., Elkind 1964; Young, 1973). A very useful, up-to-date discussion (which includes more advanced aspects of the model, including multiloop analysis) is the recent monograph of McRuer and Krendel (1974). This monograph also contains a bibliography of applications that is quite extensive.

Quasilinear models concentrate on human control strategies and tend to deal with information processing behavior in an explicit fashion only when overt visual scanning is an integral part of the task. The quasilinear model for human control performance consists of a describing function that accounts for the portion of the human controller's output that is linearly related to his input and a "remnant" term that represents the difference between the output of the describing function and that of the human controller.

The linear describing function portion of the quasilinear model takes on various forms depending on the precision with which one attempts to reproduce the human controller's characteristics. A fairly large body of data can be accounted for by a model of the form (McRuer, Graham et al., 1965)

$$Y_p(j\ ) = K_p \underbrace{\frac{j\omega T_L + 1}{j\omega T_I + 1}}_{\text{Equalization}} \cdot \underbrace{\frac{e^{\,j\omega\tau}}{j\omega T_N + 1}}_{\text{Limitations}} \qquad (1)$$

This describing function is composed of factors related to some human limitations, namely, reaction delays (τ) and lags attributed to the neuromuscular system (T_N) and of factors used to model the human's adaptive equalization characteristics.

The most important, and perhaps most elegant, result of quasilinear manual control theory is embodied in the "crossover model" (See Abstract No.5) which relates $Y_p(j\omega)$, to the transfer function of the controlled element, $Y_c(j\omega)$, by the equation

$$Y_p(j\omega)\ Y_c(j\omega) \doteq \frac{\omega_c\ e^{-j\omega\tau_e}}{j\omega} \qquad (2)$$

where ω_c is the crossover frequency and τ_e an effective time delay. The crossover model is a mathematical statement of the observation that human controllers choose their equalization characteristics so that the closed-loop system dynamics approximate those of a "good" feedback control system. The parameters of the crossover model, ω_c and τ_e, or those of equation (1), are generally task dependent; values for them have been traditionally selected on the basis of verbal adjustment rules that are, in turn, based on considerable experimental data. The crossover model is intended to be accurate only at frequencies in the vicinity of ω_c. However, this is often all that is necessary to adequately specify the closed-loop dynamic characteristics of a single-loop system. On the other hand, it may not be sufficient to predict detailed system performance, particularly if there is a substantial remnant component in the controller's response.

Thus, the crossover model codifies a large body of data, is easily understood, and permits reasonably complex single-loop systems to be analyzed simply (sometimes with just paper and pencil). A limitation of the model is the task dependence of its basic parameters and the requirement for verbal adjustment rules for their selection. These adjustment rules work well enough within the domain in which they were developed and empirically validated, but difficulties arise when one attempts to predict results in new situations. The problem would be more manageable if each new variable considered had an effect on the model parameters that could be predicted independently of other variable changes, but this is not usually the case. Thus, unless a particular combination of conditions has been studied empirically, the basis for *predicting* results for these conditions is tenuous; one is forced either to develop a catalog of all possible combinations of conditions or to invent a model, one level removed, that provides formal rules for setting parameters.

The describing function approach to multi-input, multi-output manual control systems is based on the multi-loop analysis techniques of classical control theory and on the methods and insights developed for the single-loop systems. Unfortunately, the attempt to extend single-loop results to multi-loop manual control problems turns out to be an enormous step for a number of reasons. First, there is the problem of predicting what loops the human will close. There are always alternatives and, generally, the structure cannot be "identified" uniquely from data in multi-loop situations (see, e.g., Stapleford, Craig and Tenant, 1969).* Therefore, it is necessary to postulate loop

*What data there are for multi-loop describing functions (Stapleford, McRuer and Magdaleno, 1967; Stapleford, Craig and Tenant, 1969; Weir and McRuer, 1972) indicate that, in general, single-loop rules seem to apply for "outer loops" but not necessarily for "inner loops."

structures and, often, to employ fixed-form models in each loop. Of course, each describing function in each loop will have adjustable parameters that must be specified – and, as in the single-loop case, they are task-dependent. Besides being non-unique for fixed conditions, loop closures will also depend on display and control system design in a complicated fashion. The determination of the multiplicity of pilot parameters must be done iteratively. Thus, the simplicity and ease of calculation associated with the single-loop crossover model evaporates, and one is confronted with substantial problems with respect to both prediction and computation.

The situation with respect to the remnant portion of the quasilinear models is less well developed. The current view of remnant in quasilinear manual control theory is that, in the absence of display scanning, remnant is due largely to irreducible stochastic variation in the human operator* (McRuer and Krendel, 1974, p. 65). Remnant is not error in modeling the deterministic portion of the controller's response, although such errors could contribute to remnant. Models for single-loop remnant consist of empirically obtained first-order noise spectra injected at the operator's input (McRuer and Krendel, 1974, p. 34). Fairly elaborate models for multi-display scanning have been developed and have been used to predict remnant in multi-loop situations (see, e.g., Allen, Clement and Jex, 1970). However, the scanning model does not appear to yield entirely adequate results in predicting remnant (Machuca and Lind, 1971) and, because of the interactions between loop closures, scanning strategies, etc., appears to be difficult to use.

*This interpretation is made explicit in the optimal control model and leads to a method for predicting remnant (see below).

Finally, we should mention the fact that describing function models are frequency-domain models whose data base is strictly valid only for stationary, random inputs. The describing function is not the pilot's transfer function. There have been efforts to develop so-called dual-channel controllers that will allow the quasilinear approach to be extended to problems involving step inputs, pursuit tracking, quasi-predictable inputs, etc., and these have had some limited success (see McRuer and Krendel, 1974, p. 48).

2.3.1.2 Fixed Form, Parameter Optimization Models

In an attempt to overcome some of the problems associated with selection of parameters of describing function models according to verbal and empirical adjustment rules, some investigators have used fixed-form models (FFM) with the parameters of the model selected so as to optimize some criterion. Because the fixed-form models have the same or similar forms as the observed describing-function models, they are generally considered to be of the same class of models. However, this view is not strictly correct and is responsible for some confusion with respect to the interpretation of remnant; in the FFM case, the spectrum of the modeling error is sometimes considered to be the human operator's remnant.

The most interesting of the FFM approaches was first proposed by Anderson (1970), and has come to be called Paper Pilot (see Abstract No. 6) after the computer implementation of the method (Dillow, 1971). The distinctive feature of Paper Pilot is the assumption that the human pilot chooses an equalization strategy that minimizes a cost expression that incorporates quantities related to both performance and workload,

where workload is defined in terms of the pilot lead time constants in the various loops (an interpretation that is based on describing function model results). This expression is then identified with the subjective "rating" of vehicle handling qualities provided by the human pilot. The amount that each performance and workload term can contribute to overall rating is constrained. This prompts the interpretation, given by Dillow (1969), that the pilot adapts his parameters to minimize a linear combination of workload and performance, and "washes out" exceedingly poor or good performance in determining the rating.

Anderson initially applied his approach to data obtained by Miller and Vinje (1968) from a simulated VTOL hover task. These data include pilot ratings, performance scores, and estimates of pilot parameters (obtained by adjusting the parameters of a fixed-form model to match the observed scores). Anderson used regression analysis to determine an expression that related the observed pilot ratings to a combination of measured performance scores and derived pilot (lead) parameters. He then postulated a fixed-form model and found the parameters of the model that optimized this expression. The resulting "optimal" FFM was used to "predict" scores and pilot ratings. This resulted in excellent agreement with respect to pilot rating,* but the correspondence between predicted scores and measured scores was not nearly as good. Of course, the disagreement in scores was reflected in a disagreement between predicted and "measured" pilot parameters. This inconsistency suggests that matching pilot rating (to within $\pm$ 1 unit) is not sufficient to uniquely determine the pilot model

*Given the variability of pilot rating data, a "good" match corresponds to predicting a rating that is within one unit of the rating determined experimentally, where pilot rating is measured on a Cooper-Harper scale of one-to-ten (Harper and Cooper, 1966).

parameters. Since Anderson's original work, Paper Pilot has been applied to a variety of other tasks to yield results of some promise (see, e.g., Anderson, Connors, Dillow, 1971; Arnold, Johnson, Dillow, 1973; Stone, 1973).

Thus, FFMs with model parameters selected to optimize some cost function have been applied with some success. This approach replaces the verbal adjustment rules of describing function theory with a systematic procedure that should be inherently more "predictive." However, there are serious problems and limitations associated with the approach that have yet to be resolved. In the case of Paper Pilot, schemes for a priori selection of a rating function are needed if the method is to be truly predictive. Moreover, the technique has been a less successful predictor of performance and pilot parameters than of ratings. Part of the difficulty in predicting performance (and, to a lesser extent, parameters) is the lack of a remnant model. It may be more difficult to get good remnant predictions for FFMs because of the confounding of modeling errors with remnant.

Perhaps the most serious drawbacks to these techniques arise when addressing multi-output, multi-axis, multi-control problems. Then, as in the describing function approach, it is necessary to postulate possible loop structures and the models for each loop. This complicates the problem of choosing a rating-cost function, increases the possibilities of modeling error, and increases the number of parameters to be optimized. These factors undoubtedly jeopardize the predictive capability of the techniques and also magnify the computational problems significantly.

The problems associated with computing the optimizing parameters are far from trivial. For example, gradient-type schemes, besides being slow in convergence, often converge to local minima. This problem increases in severity with increases in the number of parameters. In addition, bounds may be required on the variable parameters so as not to violate human response capability and to assure physically meaningful values (e.g., leads must be positive). When constraints are placed on the independent variables, the numerical optimization methods (gradient projection) become less efficient. Finally, the computation time requirements to attain a maximum (or minimum) increase approximately as the number of parameters squared. Thus, for multi-loop systems — where each loop can contain several parameters — computer time can soon become excessive.

2.3.1.3 The Optimal Control Model of the Human Operator

A second control-theoretic approach to human controller modeling has emerged. This approach is based on modern control and estimation theory. The resulting model has been referred to as the optimal control model (OCM) of the human operator (see Abstract No. 7), although many of the key features of the model deal with the human's information processing capabilities and behavior.

The human controller is self-adaptive and, if motivated and given information about his performance, will attempt to change characteristics so as to perform better. On the other hand, human performance is limited by certain inherent constraints or limitations and by the extent to which the human understands the objectives of the task. These observations serve as the basis for the fundamental assumption underlying the OCM, namely, that the well-motivated, well-trained human operator will act in a near

optimal manner subject to the operator's internal limitations and understanding of the task. This assumption is not new in manual control (e.g., Roig, 1962; Leonard, 1960; Obermayer and Muckler; 1965) or in traditional human engineering (e.g., Simon, 1957, calls it the Principle of Bounded Rationality). The novel aspects are the methods used to represent human limitations, the inclusion in the model of elements that compensate optimally for these limitations and the extensive use of state-space concepts and the techniques of modern control theory.

Clearly, if the basic optimality assumption is to yield good results, it is necessary to have reliable, accurate, and meaningful models for human limitations. Insofar as possible, these models (or their parameters) should reflect intrinsic human limitations or should depend primarily on the interaction of the operator with the environment and not on the specifics of the control task. It is also desirable that the description of human limitations be parsimonious and that it be commensurate with the modern control system framework that is being employed. These principles have guided the development of the models for human limitations that will be described below.

There are several reasons for choosing a modern control approach to modeling the human controller, even though methods based on classical control theory have been fairly successful. A principal motivation is the basic logic of the optimality assumption. The basic approach to human limitations and the optimality assumption suggest a model that might adapt to task specifications and requirements "automatically" and not through a subsidiary set of adjustment rules. Further, state-space techniques promise a systematic approach to multi-input,

multi-output systems that avoids some of the difficulties associated with the application of multi-loop analysis to man-in-the-loop problems. The powerful computational schemes associated with these techniques are also attractive in light of complex monitoring and control problems that are of interest. Finally, the use of a normative model* and time-domain analysis should facilitate "modular" and "graceful" development of the model as new facets of human behavior are considered and understood.

The resulting model is a stochastic, time-domain model for the human. It includes a model for predicting the random component of human response and is not limited to stationary control situations. The basic model is composed of the following: (1) an "equivalent" perceptual model that translates displayed variables into noisy, delayed, perceived variables; (2) an information processing model consisting of an optimal estimator and a predictor that generate minimum variance estimates of the system state from the data concerning perceived variables; (3) a set of "optimal gains" chosen to minimize a quadratic cost functional (a generalization of the mean-squared error criterion that expresses task requirements); and (4) an equivalent "motor" or output model that accounts for "bandwidth" limitations (frequently associated with neuromotor dynamics) of the human and his inability to generate noise-free control inputs.

The time-delay, observation- and motor-noises, and the neuromotor-lag matrix account for inherent limitations on human processing and perceptual-motor activity. Methods for choosing values for these quantities have been determined by matching

*The model is normative in that it predicts what the human *should* do, given his limitations and the task. Thus, for a new situation one need only determine the operative limitations and what should be done. The fact that this assumption works well is testimony to the adaptability and capability of the trained human operator.

experimental data and these values have been found to be generally independent of task parameters, (Kleinman, Baron and Levison, (1970). The observation noise is a key feature of the model. It is, essentially, a lumped representation of human randomness. From the standpoint of classical quasilinear describing function theory, the observation noise may be thought of as a model for controller remnant. On the basis of considerable experimentation, a relatively simple set of rules for predicting remnant has been found (Levison, Baron and Kleinman, 1969).

The optimal estimator, predictor, and gain matrix represent the set of "adjustments" or "adaptations" by which the human attempts to optimize his behavior. The general expressions for these model elements depend on the system and task and are determined according to well-defined rules by solving an appropriate optimization problem. The solution to the optimization problem yields predictions of the complete closed-loop performance statistics of the system. Probability densities of all system variables (states, outputs, and controls) are generated as functions of time, along with mean and rms error deviations from a nominal path. Moreover, the densities of the operator's estimates and estimation errors are also predicted as functions of time. All computations are performed using covariance propagation methods, thus avoiding costly Monte Carlo simulations. However, if desired, a "sample" or simulation version of the model is possible (see below).

The optimal control model has been subjected to extensive validation with extremely encouraging results. It has been validated in relatively simple, stationary control tasks (Kleinman, Baron and Levison, 1970) and in more complex tasks, both stationary (Baron et al., 1970) and non-stationary (Kleinman and Baron, 1973; Kleinman and Perkins, 1974; Kleinman and Killingsworth, 1974; Baron and Levison, 1974).

The optimal control model of the human operator appears to have several advantages. First of all, there has been considerable success in isolating human limitations so that they are essentially task independent. That is, the parameters of this model that describe human limitations do not vary significantly with vehicle dynamics and forcing function bandwidths so long as the underlying linearizing assumptions remain valid and so long as the operator is motivated to perform at his limits. In those cases where performance does not degrade substantially when the operator does not devote his full energies to the task, the parameter values adopted by an operator will in all likelihood depend on subjective factors (such as psychological set); however, in these cases performance predictions using the model will remain reasonable and the model itself can be used to determine an appropriate range of operator parameters (Levison, Elkind and Ward, 1971; Levison, Baron and Junker, 1976).

When the characteristics of the display or control manipulator are changed, the approach in the OCM is to model these changes directly, where possible, rather than to adjust parameters of the model according to some other rules. This has been accomplished in the case of visual thresholds (Kleinman and Baron, 1973) and, more recently, in the case of control-manipulator dynamics (Levison and Houck, 1975). The resulting changes in the pilot's describing function and remnant are then predicted as an output of the OCM. This constitutes a fundamental difference from the describing function approach wherein the resulting changes in describing function and remnant are generally catalogued (see, e.g., Jex, Allen and Magdaleno, 1971).

In the OCM, there is no need to postulate loop closures or model forms for the various loops. Instead, one must specify what is displayed to the pilot (this is generally given) and the cost function to be optimized (see below). The OCM adapts to changes in system dynamics and forcing function changes via the solution to a well-defined optimization problem that admits a closed-form solution.

Another feature of the OCM is that a simple and effective model for remnant is incorporated. This has permitted accurate predictions of remnant and, therefore, performance. Moreover, the interference model (Levison et al., 1971) allows for accurate remnant prediction in multi-axis tasks that do not involve overt scanning; because of the close connection between attention-sharing and scanning, it is possible that scanning remnant can also be predicted accurately.

The problem mentioned most often in connection with the OCM is the selection of a cost functional. In applying the model to predict performance, it is assumed that the human operator will perform the control task so as to optimize a quadratic functional of the system variables. This functional is defined objectively by specifying relative penalties for various system errors; this specification involves a careful analysis of the problem. However, one cannot be certain that human controllers will optimize the objective functional rather than some subjective criterion. Initially, it was believed that this was an area where artistry in applying the model was needed (Young, 1973) but experience with the model has revealed few, if any, instances where the choice of performance criteria has actually posed serious difficulties. Nevertheless, the specification of the cost functional remains the primary area involving the judgement of the system analyst in applying the OCM.

The information processing portion of the OCM includes an estimator and a predictor. These elements contain perfect "internal" models of the system. While the notion that the trained human operator has an internal model is quite appealing, the assumption that the model is perfect appears to be less well justified. To some extent, model imperfections are accounted for by the observation and (especially) motor noises. Moreover, results to date have indicated that so long as there are random disturbances of any significance, errors introduced by the assumption of a perfect model tend to be negligible. In other circumstances, however, the perfect model assumption may lead to inaccuracies in model predictions.

A related problem concerns the increased order of the pilot model with increases in problem complexity. Modern control theory (and the OCM approach) treat input shaping-, sensor-, actuator- or display-dynamics, by state augmentation. The OCM will have a perfect representation for these dynamics and will increase in order accordingly. This increases the computational expense for obtaining a solution and suggests the possible desirability of some reduced-order modeling.

Another problem concerns the motor portion of the model. This model may need further refinement, especially with respect to multi-axis control tasks.* Presently, there is no accounting for inadvertent control crossfeeds. In addition, the motor noise/signal ratio has been used largely as a surrogate for internal model imperfections (Baron, et al., 1970). A residual motor noise, that doesn't scale with the control command, should be incorporated to account for possible motor thresholds and for the experimental evidence showing that controllers will introduce noise into an undisturbed system.

*Until very recently, the tendency has been to concentrate on display-related problems.

Finally, rigorous schemes to estimate and establish confidence limits for the underlying model parameters of the OCM do not presently exist. Attempts to identify the parameters have run into some difficulty and it has been claimed that the parameters of the OCM are not uniquely identifiable (Phatak, Weinert and Segall, 1974). It is clear that some problems can arise in formal procedures because what is optimal with respect to a particular cost functional and set of parameters is apt also to be optimal with respect to a slightly different cost functional and parameter set. On the other hand, the heuristic methods that have been used have yielded parameter estimates that predict new results successfully.

2.3.1.4 Simulation Models

As our definition, we take a manual control simulation model to be a model which produces a specific time-history for the human controller's output given a specific time history for the input.* All manual control models describe the input/output characteristics of the human operator in a manner that is consistent with an instant-by-instant description. However, the models that have dominated the field (see previous sections) have generally avoided using their descriptions to predict instantaneous response of the operator; rather, the emphasis has been on statistical predictions of performance and determination of stability and vehicle handling quality via closed-loop linear analysis techniques. There are several reasons for this emphasis. Most importantly, the capacity for generalization and, hence, prediction is much greater when employing the theoretic-analytic models described above. Other reasons include the fact that many problems of interest involve response to random inputs, thus requiring a statistical

*We do not mean time-histories of the ensemble statistics as produced, for example, by the OCM. However, a sample path from such an ensemble would qualify as the output of a simulation model.

analysis; that inter-subject differences, as well as the stochastic behavior within individuals, makes prediction of a particular response essentially impossible; the computational efficiency of linear analysis techniques; and, finally, that simulation models frequently degenerate into a "fiddler's paradise."

On the other hand, there are some strong motivations for simulation modeling. If one wants to integrate a manual control model with another model this will often require a simulation representation. Certain types of system or operator nonlinearities may not be amenable to linearization techniques and will require simulation for analysis. Statistical analyses can sometimes obscure certain important response patterns. Finally, analyzing individual responses may ease the interpretation of results and the validation of models.

There have been many simulation models developed and reported in the manual control literature, often differing from one another in a minor fashion. As might be expected, this development reflects the developments in linear, analytic models (which, in turn, reflect advances in control theory). Thus, early simulation models include transfer functions that frequently derive from quasilinear models while later ones are reminiscent of FFM's or the OCM. Here, to give a flavor of the approaches to simulation modeling, we shall discuss some representative models.

One of the first attempts at simulation modelling of human pilots was conducted by Diamantides (1958). The simulation model was constructed from analog computer elements and was composed of elements thought to be representative of human linear and nonlinear behavior. Thus, the continuous model incorporated neuromuscular dynamics, reaction delays and equalization

transfer functions similar to those of quasilinear models. It also included direct perception of derivatives of displayed signals and a number of nonlinearities (perceptual and indifference thresholds, anticipation and dither circuits, and saturation elements). The parameters of the "analog pilot" were varied until the simulation model exhibited responses that were quite similar to those of actual pilots. The model proved to be interesting and instructive. It reflected what was known from quasilinear theory and the prevailing technology of using analog computers to analyze complex, nonlinear control systems.

Costello (1968) developed the so-called "surge model" to predict operator response to high bandwidth inputs. The surge model has two modes: for "sufficiently small" error and error-rate, a linearized, constant coefficient model is used; when the error and error-rate exceed pre-determined bounds, and whenever the phase-plane trajectory crosses a switching line, a pre-programmed double step or "surge" is introduced. The response of the "surge" model to step inputs, square wave inputs and chopped sine wave inputs compares favorably with responses to these inputs observed experimentally, more so than the response of the simple linear, constant coefficient portion alone. Costello's model is one of many simulation models (e.g., Johanssen, 1972, see Abstract No. 8; Phatak and Bekey, 1969, see Abstract No. 9; Veldhuysen, 1976, see Abstract No. 10: that employ "decision surfaces" in the phase-plane to change modes of operation.

A recent, ambitious development in simulation modeling was undertaken by Onstott (1974, see Abstract No. 11). His model incorporates fixed form pilot models. An "urgency function" is defined to determine when the model should "switch attention" from one axis to another. An attempt is also made to account for remnant in the model, including the effects of inadvertent control

crossfeeds. Onstott claims close agreement between model and experimental data, though the results given in the above cited reference appear somewhat mixed. Again, this model reflects the trends in analytic modeling, particularly the use of fixed-form pilot models with parameters selected to optimize some performance criterion.

We close this discussion by reemphasizing the point that the analytic models, though not developed for simulation purposes, can serve as the basis for simulation models. We have seen that many simulation models employ transfer functions that are similar to the describing functions developed in quasilinear theory. (This is done because the form of the describing function is reasonable and despite the fact that, from a theoretical standpoint, the describing functions are not strictly valid for this purpose). It should also be mentioned that the Optimal Control Model, described earlier, can be used as a simulation model (Kleinman and Perkins, 1974). The simulation version of the OCM requires selection of a particular realization of the various sources of problem randomness (both system and human) to generate the simulation "histories." Since the model is a time-domain model and admits time-varying parameters, it can be quite general in its simulation extension while retaining features verified earlier. On the other hand, a generalized simulation version of the OCM has not been validated experimentally.

2.3.2 Models for Visual Scanning

In the manual control field much of the impetus for analyzing and modeling visual scanning behavior has been provided by the aircraft control problem. A large, but frequently deficient, data base exists with respect to monitoring aircraft instrument

panels (see Barnes, 1972, for an analysis of a significant portion of aircraft eye-movement work). Several models for predicting and/or describing scanning behavior have been suggested. These models have their roots in information theory and in statistical estimation and decision theory. All the models we discuss here are relevant to pure monitoring problems. They differ in their theoretical bases and in the degree to which they consider the effects of closed-loop control on scanning.

The first quantitative model for describing pilot sampling behavior was developed by Senders, (1964, Abstract No. 12). This model was based on information-theoretic ideas, particularly Shannon's sampling theorem. A basic assumption of the model was that the human observer samples the various signals periodically and attempts to reconstruct the time functions presented on each instrument. Moreover, it was assumed that the operator was effectively a single channel device capable of attending to only one signal at a time. With these concepts as a starting point, Senders was able to derive expressions for the frequency and duration of samples of an instrument given its input signal characteristics and the required precision of readout. This model predicted quite well the average of subjects in an experimental situation, but the agreement obtained was somewhat fortuitous in that it depended on a unique experimental condition (Senders, et al., 1969).

However, it seemed reasonably clear that the simple periodic sampling model would not adequately predict behavior in more complex situations, especially since observed data gave evidence of aperiodic sampling behavior. Taking a cue from the fact that pilots often are only concerned with detection of extreme readings rather than with

signal reconstruction, Senders, et al. (1966) proposed a conditional sampling scheme that would result in aperiodic behavior. In this approach, the human monitor is considered as a channel for the transmission of discrete messages and not as a channel for the transmission of a complete time function. Then, it is possible to postulate several, not necessarily mutually exclusive, sampling strategies (e.g., sample when the probability that the signal exceeds a prescribed limit is greater than some subjective probability threshold or when the probability of exceeding the limit is a maximum) (Senders, et al., 1966).

Smallwood (1966) (Abstract No. 13) developed a model which, although similar in some of the details, was conceptually quite different from those previously developed. His model was based on two underlying assumptions: (1) the human operator bases his state of information about his environment upon an internal model of this environment, the model being formed as a result of past perceptions of his environment;(2) the human operator behaves optimally with respect to his task and his current state of information within his physical limitations. To apply the model, Smallwood makes further assumptions. First, he postulates a form for the internal model that describes the monitor's conception of the environment he is monitoring. Smallwood's approach to this problem is to assume that the monitor's model of each instrument is a good approximation to the true situation. (Note the resemblance to the Kalman filtering approach.) Smallwood's interpretation of optimality was that the human monitor, interested in detecting immediate excursions of the instruments beyond the threshold, switches his attention to that instrument for which the probability of exceeding the threshold is a maximum. Thus, he dealt strictly with monitoring, essentially ignoring the control objectives.

The above-mentioned models have several limitations that arise either from inherent factors or, in the case of Smallwood's model, from a failure to exploit the full potential of the approach. One significant limitation was that none of the models accounted for correlation between signals on various indicators. It is clear that with an appropriate choice for an internal model, a method for dealing with coupling is possible within the framework of Smallwood's model, but he did not attempt to include such effects. A second significant limitation of these models is that none of them take explicitly into account the interaction between sampling and control behavior. In all cases the human operator is considered as a monitor only; the control requirements play little, if any, role in the selection of a sampling strategy. It is true that *a posteriori* signal analysis, which is necessary to obtain the parameters of the conditional sampling models, will include the effects of control. However, this is quite different from using the control task in an attempt to predict sampling behavior.

All of the models consider the human operator as a single-channel device capable of processing information from only one instrument at a time. This approach is at variance with the experimental data obtained by Levison and Elkind (1967), concerning peripheral tracking. Their experiments showed that under certain conditions, pilots could perform two-axis compensatory tracking even though the information concerning one of the axes was always in the peripheral visual field. While there might be some argument as to whether the pilot is actually processing both signals simultaneously, it seems fairly clear that the ability to perform peripheral tracking will in fact affect the pilot's sampling behavior and should therefore be accounted for in a sampling model. Still another limitation of the models is

that no risk or cost structure has been incorporated. Equal costs are assigned to all instruments. Moreover, there is no cost assigned to taking a sample or to switching attention.

Carbonell (1966, Abstract No. 14) attempted to overcome some of the above limitations by developing a model of visual sampling which had its roots in queuing theory. If the human is assumed to be a single-channel processor, then one is led to the notion that the various information sources, i.e., instruments, queue up and wait their turn to be processed. The analysis of scanning can then be approached as a problem in queuing theory and one can arrive at estimates of the probability distribution of simultaneous demands, the probability distribution of waiting times,and estimates of the probability that events of interest will be missed.

Carbonell's model is clearly more general and more flexible than the models previously discussed. In addition, it has achieved good accuracy in a validation study of approach to landing. Nevertheless, the model has not removed all the previously cited limitations. Peripheral processing is not accounted for; indeed, the concept of the human operator being a single-channel processor is central to the idea of the instrument queue. Also, coupling among instruments was not included in the model, although Carbonell claimed that such coupling could be incorporated. Finally, it should be pointed out that the model incurs a heavy price in analytic complexity for the flexibility that is obtained. It appears that only through extensive simulation can one obtain the pertinent model parameters and predict human sampling behavior.

Allen, Clement, and Jex (1970) (see Abstract No. 15) and McRuer et al. (1968) have attempted to synthesize the concepts of Senders' scanning model, multi-loop describing function theory, and some ideas of Clement concerning human signal reconstruction into a theory for displays in manual control. Allen, Clement, and Jex suggest a model for scanning, sampling, and reconstruction that comprises: (1) a quasilinear, random-input "perceptual describing function," which multiplies the human operator's continuous describing function (i.e., is inserted serially in the control-loop); and (2) a broadband sampling remnant, which adds to the basic remnant, and is described as a wide-band observation noise injected at the pilot's perceptual input. The "perceptual describing function" is usually specified by an attenuation factor K_h and an equivalent sampling delay (T_s). The values of K_h and T_s depend on scanning interval, dwell fraction, weighting of rate information, and type of signal reconstruction used.

The approach of Allen, Clement, and Jex ameliorates the major difficulty associated with "open-loop" scanning models in that control requirements are taken into account, albeit in a qualitative way. In addition, the approach has been partially validated in extensive simulation studies. However, the scanning model accentuates the problems already noted for the describing function approach to pilot modeling. One must face the problems of selecting multi-loop structures via the "adaptive feedback selection hypothesis" and also the problem of deciding among various signal reconstruction schemes. Moreover, the whole process is iterative in that scanning predictions must be made after the control structure is defined, but the scanning strategy can affect the control structure. Hoffman, Clement, and Blodgett (1973) have developed a technique that yields a non-iterative procedure for the case where there is a single primary instrument, i.e., a flight director, needed for control. In general, however,

the approach could remain extremely difficult to use and computationally inefficient. Another important problem is that the model for scanning remnant does not appear to yield entirely adequate results (Machuca and Lind, 1971). Accurate remnant representation is, of course, very important for predicting performance with scanning because this is invariably a high remnant situation.

Baron and Kleinman (1968) have proposed a scanning model (Abstract No. 16) for incorporation in the OCM and have modified it and applied it in a VTOL hover task (Baron, Kleinman,et al., 1970). The basic assumption in this model is that the pilot chooses his control input and his scanning strategy to minimize the quadratic cost functional describing his task requirements. The scanning strategy is defined by a set of parameters (average dwell time, scan frequency, etc.) which in turn are adjusted to minimize closed-loop cost. This model is therefore suitable for prediction.

There are several advantages of this approach to scanning prediction. The sampling model is such that the human's monitoring behavior depends upon the control requirements and control actions in an explicit way. Control and scanning strategies are determined at essentially the same time; one avoids the inherently iterative and judgmental procedures involved in a process that requires the specific loop topology to be known before computing scanning parameters. Also, one avoids separate assumptions concerning reconstruction of the sampled signal; the Kalman estimator performs this function, as in the no-scanning case. Finally, the observed effects of scanning are a natural consequence of this approach. Increased average observation noises result from the impossibility of fixating foveally all displays at once. This leads to increases in remnant, accompanied by reductions in pilot gain and an apparent increase in time delay.

The principal difficulty associated with the OCM approach would appear to be the computational burden in solving the optimal sampling problem. In the two-display case, numerical search techniques are adequate and not too expensive. For more than two displays, a suitable optimization algorithm is necessary and the analysis is likely to be expensive, but, perhaps, not more so than other approaches.

A final comment concerning scanning models seems appropriate. Experience seems to indicate that choice of a scanning strategy is highly idiosyncratic and that there is great variability in scan patterns among subjects. For example, Machuca and Lund (1971), in describing the results of a study to verify the model of Allen, Clement, and Jex, state the following:

> "The pilot-subjects adopted different average scanning and sampling strategies. Subject preferences in display orientation, control sticks, and simulation techniques were noted to influence scanning behavior."

and

> "A subject may adjust his scanning strategy (i.e., sampling interval and dwell fraction) to maintain acceptable control over the control tasks. During this experiment, scanning parameters, which attempt to account for changes in scanning performance, did not yield the predicted range of values when the controlled dynamics were changed and when simulation techniques were altered from those used by other experimenters. However, in spite of such changes, closed-loop responses remained essentially unchanged.

The latter statement indicates that the performance was not very sensitive to changes in scanning strategy over a reasonable range of variation. This is probably true for most reasonably designed displays, as indeed, it should be. The insensitivity also accounts

for the variability among subjects; there is not a great deal to be gained by adopting the "precisely" optimal scanning strategy. Thus, if one is interested in predicting performance, a relatively crude model of scanning may be adequate.

2.3.3 Manual Control Models for Workload

There has been a substantial concern with problems of operator workload in manual control. Much of this effort has been motivated by research in aircraft handling qualities and pilot opinion. However, a significant portion of the research is relevant to basic questions of task interference and the effects of attention-sharing on performance.

Models for predicting operator performance and workload must take into account the effects of interactions among several control and/or monitoring tasks. When overt visual sampling is involved, it is clear that there will be degradation that should be accounted for in a model for the operator. The question of scanning workload has been treated by Senders (1964), who derived an expression for the fraction of time that must be devoted to a given display given the amplitude and bandwidth of the displayed signal and the permissable rms reading error. With this relationship, the total workload placed on an ideal observer by a given set of displays can be computed. The extension of these ideas to problems involving closed-loop control has been carried out by McRuer et al. (1968) and their colleagues (see Section 2.3.2, above). This provides a mechanism for treating scanning workload in the context of describing function models. The attendant drawbacks of the approach have been discussed previously.

Task interference and/or attention sharing are not limited to situations involving overt scanning. For example, Brainerd et al. (1962) found that reaction time increased on the average as the number of stimulus-response channels increased. Since in his experiments all the visual stimuli could be monitored simultaneously and all motor responses executed simultaneously, no overt sampling was required. There was, therefore, interference at the central processing level as the number of tasks was increased.

Numerous studies of tracking in multi-variable situations (e.g., Chernikoff et al., 1959; Todosiev, et al., 1965; Levison and Elkind, 1967), generally confirm that performance in a given axis is degraded when an additional task is added. The amount of degradation will be very dependent on the experimental situation and in some cases may not even be significant (e.g., Todosiev et al, 1965). In addition, changes in describing functions have been observed, though these changes, too, depend on the particular characteristics of the experiment. A third effect is that as displays become less integrated and/or more separated, performance degradations increase, and are worst when overt scanning is necessary.

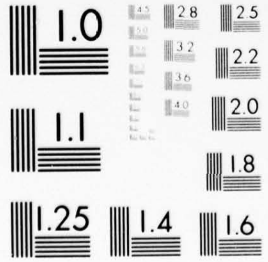
Although there is considerable concern with workload in the describing function literature, we are unaware of any theory for dealing with task interference that does not result from overt visual scanning. In single-loop problems, workload is related to the amount of lead the pilot must generate – but what happens to lead generation as tasks are added is not predicted. (Generally, effective time delays increase with task complexity.) Paper Pilot predicts requisite leads and combines them in a rating expression (Anderson, 1970). As the

AD-A038 597 BOLT BERANEK AND NEWMAN INC CAMBRIDGE MASS F/G 5/8
CRITICAL REVIEW AND ANALYSIS OF PERFORMANCE MODELS APPLICABLE T--ETC(U)
MAR 77 R W PEW, C E FEEHRER, S BARON F44620-76-C-0029
UNCLASSIFIED BBN-3446 AFOSR-TR-77-0520 NL

2 OF 4
ADA038597

2 OF 4

ADA038597



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

number of loops increase, this approach is likely to run into additional problems unless methods for considering interference in this context are developed.

Levison, et al. (1971, Abstract No. 17) have proposed a model for task interference and attentional workload that is appropriate to the optimal control model. The workload model, in conjunction with the OCM, allows one to predict how the operator will allocate his available attentional capacity among tasks. It also allows one to determine the relative attentional workload for a given task. The model is intended to apply to situations (such as continuous manual control) in which the human is required to operate as a continuous processor of information. Very briefly, the model for interference assumes that as the human devotes less attention to a given displayed variable, the noise on that variable increases proportionally. Thus, if P_o is the observation noise/signal ratio when "full attention" is devoted to a task, then when the subject is forced to pay less than full attention to the i^{th} task, the effective noise/signal ratio is

$$P_i = \frac{P_o}{f_i}$$

where f_i is the fraction of attention devoted to the i^{th} task $0 \le f_i \le 1$, and $\Sigma_i \; f_i = 1$. Clearly, because of the increase in noise, attention sharing will degrade performance.

The above equation can be used to help predict the effects of a given allocation of attention on performance. In particular, the equation serves to determine the observation noise covariances associated with that allocation (specific values for f_i and P_o). Then the optimal-control model is used to predict pilot behavior and overall system performance. If it is not known beforehand how

the operator will allocate his attention, model solutions are obtained to predict the optimum allocation of attention along with other measures of interest. In keeping with the fundamental optimality hypothesis, the optimum allocation of attention may be taken as a prediction of the operator's allocation.

Building on the model for attention, a "workload index" can be defined as the fraction of attention required to achieve a specified criterion level of performance on the control task. Thus,

$$\text{workload index} = f_{t_c}$$

where f_{t_c} is the minimum fraction of attention for which performance can be maintained within the criterion level. In order to predict the workload index, it is necessary to specify a relevant performance measure, the required level of performance, and the "reference" noise/signal ratio P_o. Levison lets P_o correspond to the noise/signal ratio obtained in a standardized laboratory situation in which the operator is motivated to minimize his tracking errors. Although this value does not necessarily correspond to "full capacity," it does appear to correspond to a high workload condition, and operation at this level for any prolonged time would undoubtedly be unacceptable. Of course, when one is interested primarily in the relative change in workload requirements from one situation to the next, the value for P_o is not too critical.

Levison's model of attention-sharing has been validated for multivariable manual control tasks (Levison, et al., 1971) and for two decision tasks (Levison and Tanner, 1971). For the control task, where detailed measurements of human response were available,

it was found that operator remnant and describing function changes and system error scores were affected by multiple-task requirements in the manner predicted by the model. It was also found that overall decrements in performance were predicted better than individual task measures. This result appears to be a consequence of the insensitivity of total performance to reasonable deviations from optimality in the allocation of attention among individual tasks.

A limitation of Levison's approach is its inability to predict absolute (rather than relative) workload. As noted by Wewerinke (1974) the above model deals with selective (rather than intensive) aspects of attention. He attempted to extend Levison's model to include the intensive aspects of attention and, thereby, predict absolute workload. He related level of arousal in a task to the sensitivity of system performance with changes in attention and defined an "absolute" workload index by

$$WI = f_{t_c} \cdot S$$

where S is the sensitivity defined as the fractional change in performance for a given change in attention evaluated at the criterion level of attention. This predicted WI was compared with subjective pilot ratings of task difficulty for six control tasks of varying difficulty; there was a strong linear correlation ($r = .99$) between the two.

Finally, we should note that Levison's model applies to the prediction of scanning workload as well. For scanning, the fractions of attention sum to less than unity because of the loss of information during large eye movements. Once this lost time is accounted for, one can determine the optimal dwell fractions. However, scanning workload is not predicted by

adding dwell fractions as in Senders' models; rather, it is determined by computing f_{t_c}, subject to the modified constraints.

2.3.4 Decision-Making Models in Manual Control

In manual control, the main attempts at modeling decision-making performance have been directed at problems of failure detection and identification, unless one considers scanning as a decision-making process. (See Young (1969) for a fine review of the early work in this area). Most of the effort concerns behavior or performance following an abrupt change in system dynamics, as might occur, e.g., in a failure of an aircraft stability augmentation scheme (Elkind and Miller, 1967, Abstract No. 18; Phatak and Bekey, 1969, Abstract No. 9; Niemala and Krendel, 1974). It is generally agreed that the adaptive response of human controllers in failure situations may be considered to comprise the following basic phases (Elkind and Miller, 1967; Phatak and Bekey, 1969; Phatak and Kleinman, 1972):

(1) Control of pre-failure plant dynamics
(2) Detection of a change or failure in plant dynamics
(3) Identification of new dynamics
(4) Modification and optimization of control strategy as appropriate to new dynamics

The first and fourth phases of this adaptive response are modeled by traditional manual control models, whereas detection and identification are addressed by decision-making models.

Two fundamentally different approaches to modeling the process of detecting a failure are evident in the classical manual control literature. One approach relies on the concept of detecting patterns in the observed signals while the other employs the concept of an internal model. Frequently, the pattern recognition approach has reduced to defining decision boundaries in the phase-plane (error - error rate space): crossing appropriate boundaries in this space is taken as an indication of a change in pattern warranting a decision that a change in dynamics (or a failure) has occurred. Perhaps the most significant exemplar of this approach is provided by Phatak and Bekey (1969), who performed a detailed and interesting analysis of adaptation to stability augmentation failures in a VTOL aircraft.

The alternative approach to failure detection is based on the notion that detection occurs when an unexpected response is observed and when the operator is conditioned to accept the possibility or probability of a failure. This approach depends on an internal model for the system to generate the expected response. Several schemes of this type have been proposed (Young and Stark, 1965; Knoop and Fu, 1964; Miller, 1965; Elkind and Miller, 1967). The work of Elkind and Miller (1967) seems to be the most sophisticated and detailed of these efforts. They postulate that the operator continuously monitors the change in the system error rate that results from his control inputs, and compares this change with the predicted change that he expects on the basis of his model of the system dynamics. If the difference between the observed and expected changes in error rates, relative to the standard deviation of the distribution of expected differences, exceeds a criterion value, the model reports that a change in system dynamics has occurred.

Models for identification of the post-failure dynamics for the above approaches follow from the detection models. A common theme, however, is that the operator is familiar with (i.e., has a model of) the post-failure dynamics. Thus, in the Phatak and Bekey model, a supervisory algorithm has stored strategies, as determined from quasilinear manual control models, appropriate to each of three possible failures. The model goes through a sequential process based on observation of error and error-rate. In particular, once a failure has been detected (i.e., when the phase trajectory leaves the "unfailed region" of phase space), the strategy for one set of post-failure dynamics is tried. Based on changes in the sign of error-rate and on leaving or entering various decision regions in the phase plane, the model decides whether these dynamics are appropriate or not. It continues this process until it stabilizes on an appropriate strategy.

The Elkind and Miller model also assumes that the operator is well-trained and familiar with all possible changes in the dynamics that might occur. Once the detection portion of the model reports that a change in system dynamics has occurred, the model then enters a Bayesian analysis phase in which the controller compares the observed change in error rate with that expected under alternative hypotheses concerning the system change and selects that set of alternative dynamics that best accounts for the observed results. Usually, only one candidate will match adequately, but if not, identification is postponed and additional observation are accumulated until an unequivocal choice can be made. In addition to the distinction between what is processed to detect a failure, two other features of this identification scheme distinguish it from that of the Phatak and Bekey model:

(1) the use of statistical decision theory (hypothesis testing) for deciding among alternative dynamics, and (2) the use of a parallel rather than a sequential selection process.

The models of Phatak and Bekey and Elkind and Miller have each been compared with experimental data in a limited number of situations. These comparisons suggest at least good qualitative agreement and, in the latter case, some quantitative agreement as well. However, the models seem to be limited in their potential for application to new problems. For example, the idea of error pattern recognition is general but the mechanization of decision surfaces is clearly dependent on the specific dynamics. Phatak and Bekey do not provide any suggestions or theory as to how to choose such regions a priori (i.e., before an experiment). With respect to the Elkind and Miller model, their specific implementation would appear to have conceptual, if not practical, difficulties with detection of failures in higher order or multivariable plants (Young, 1969).

Developments in modern control and estimation theory have paved the way for extension of the above concepts to decision-making models of potentially greater generality. These developments are reflected in the information processing structure of the OCM and the resulting decision-making models may, therefore, be viewed as extensions of the OCM. However, even within this common framework there is considerable room for diversity. To see this, let us briefly review the salient features of that information processing structure.

Recall that the OCM contains an optimal estimator as part of its perceptual processor. The estimator processes the available observations to generate a minimum variance estimate of the system

state, $\hat{\underline{x}}(t)$, and the covariance matrix, $\underline{\Sigma}(t)$, of the error in that estimate. The manner in which this estimate is generated is of interest, too. In particular, an internal model of the system is used to predict an estimate of the state and of the observation. The predicted state estimate is updated or corrected using the observation "residual," $\nu(t)$, which is defined as the difference between the actual observation and the expected observation. When the estimation process is optimal, which requires that the internal model be identical to the true system model, the residuals are a white noise process (with covariance $V_y(t)$ equal to the observation noise covariance); in other words, all information about the state of the system has been extracted from the measurements.

It is apparent from the above that the optimal estimator generates variables appropriate for either a pattern recognition approach ($\hat{x}(t)$) or an internal model approach ($\nu(t)$) to detection. If just this information is used, one can develop descriptive models of detection or decision-making. For example, a purely descriptive model based on the state estimate could involve decision regions in state space just as in the pattern recognition models. Alternatively, a descriptive detection model could be based on a test on the residuals analogous to Elkind and Miller's scheme or on a test of the "whiteness" of the process. Note that the latter schemes are not tied to any particular set of dynamics.

For prescriptive models of decision-making, optimal decisions must be generated, and one cannot rely solely on $\hat{\underline{x}}(t)$ or $\underline{\nu}(t)$. However, the necessary information for making such decisions is provided by the optimal filter in either ($\hat{x}(t)$, $\Sigma(t)$) or in ($\nu(t)$, $V_y(t)$). Whether one uses ($\hat{x}, \Sigma$) or (ν, V_y)

as an input to the decision mechanism depends to some degree on the particular problem and to some degree on one's view of the operator's perceptual process. From an analytical standpoint, it would appear that the use of the residuals is preferable for several reasons (Gai and Curry, 1976). In the two applications of these techniques to be discussed shortly, Levison and Tanner (1971) use the state estimate and Gai and Curry (1976) the residual representation.

Many of the above ideas have been discussed by Phatak and Kleinman (1972). They also showed how one could employ "multiple" internal models in the OCM structure to perform simultaneous detection and identification. Their paper was theoretical and did not contain validation data. However, two studies of detection that have tested some of these ideas are now discussed.

Levison and Tanner (1971, Abstract No. 19) studied the problem of how well subjects could determine whether a signal embedded in added noise was within specified tolerances. Their experiments were a visual analog of classical signal detection experiments except that "signal-present" corresponded to the situation of the signal being within tolerance. From a practical standpoint, the decision is analogous to the problem of deciding whether conditions warrant a given action; e.g., deciding whether one is in the approach window on landing or deciding whether tracking errors are small enough for weapons release. The problem differs from those we have discussed thus far in that it is not concerned with an abrupt change in system dynamics as in a sudden failure.

Levison modeled this situation using the OCM information processor and assuming that the operator is an optimal decision maker in the sense of maximizing expected utility. For equal penalties on missed detections and false alarms, this rule reduces to one of minimizing expected decision error. The decision rule is, simply, a likelihood ratio test, which Levison implemented using the pair $(\hat{\underline{x}}(t), \underline{\Sigma}(t))$. This pair is appropriate for this problem because the decision depends wholly on the system state.

Model predictions compared favorably with experimental conditions for a variety of cases involving different signal/noise ratios and different noise bandwidth. In addition, Levison tested the model for the case of two concurrent decision tasks and a concurrent decision and manual control task. This tested the task interference model for the OCM (Levison, et al., 1971) as well as the decision making model. Good results were obtained for the two decision tasks but not for the combined control/decision task. Levison suggests a number of methodological explanations for the lack of agreement in the latter case, but this remains an area for further investigation.

Gai and Curry (1975, 1976, Abstract No. 20) have used the OCM information processing structure to analyze failure detection in a simple laboratory task and in an experiment simulating pilot monitoring of an automatic approach. In both cases the "failure" did not correspond to a change in system dynamics but, rather, was analogous to an instrument failure. To be more specific, a step or ramp was added to the observed signal at a random time. This produced a non-zero mean value for the signal and for the residual; failure detection consisted of testing an hypothesis concerning the mean of the distribution of the residuals.

Gai and Curry used sequential analysis to perform the hypothesis test. By summing (or integrating) the residuals, a likelihood ratio can be calculated and used to arrive at the decision. Gai and Curry modified classical sequential analysis to account for the fact that a failure detection problem is characterized by a transition from one mode of operation to another at a random time and the classical analysis is based on the assumption that the same mode of operation exists during the entire observation interval.

They reported good agreement between predicted and observed detection times for both the simple and more realistic situations. In the latter case, the model is used in a multi-instrument monitoring task and accounts for attention sharing (via Levison's model) and cross-checking of instruments to confirm a failure. A significant result of the experiments is that the property of integration of the residuals appeared to be confirmed for both step and ramp type failures.

The studies of Levison and Tanner and Gai and Curry suggest significant potential for the use of the modern estimation techniques in modeling decision-making. However, they remain limited in scope; for example, techniques for identifying a failure such as those proposed by Phatak and Kleinman have yet to be tested. We believe the most significant aspect of this work is that it provides verification of the information processing structure of the optimal control model in situations not involving closed-loop control. When these results are added to the implicit validation provided by the tracking data, one has a strong case for this type of modeling of human information processing. A final point worthy of note is the analogy between this approach and that of classical signal detection theory (Green and Swets, 1966).

Here, as in the classical theory, the perceptual process is corrupted by an internal (observation) noise, and, moreover, it is separated from the decision criteria. It would appear that this analogy would be fruitful to explore.

2.3.5 Summary

Manual control models are undoubtedly unique among human performance models in the degree of their quantitative description and the extent to which they have been validated and applied. The models have been employed to analyze new flight control and display problems, to determine the effects of stressors on performance, and to assist in experimental and simulation planning. It is clear that the problem of modeling single-input, single output, time-invariant systems is essentially solved. The corresponding multi-loop modeling problem, though not completely solved, is in relatively good shape too. On the other hand, models for time-varying systems, non-linear systems and for supervisory control are, comparatively, only at an early stage of development. This is true, too, of the problems associated with predicting the effects of stress. Moreover, all the models are designed to predict the performance of well trained subjects, usually drawn from a group (pilots) that have already undergone a pre-selection process. At present, very little of a quantitative nature can be said concerning performance during training.

With respect to the basic controller models that have been developed, any or all of them could prove useful in modeling aspects of the command and control situation that involve continuous control. However, the real challenge will be to see if the techniques developed for manual control can produce similar success in modeling monitoring and decision-making tasks, which usually require infrequent control action, that are typical of

large portions of the C^3 problem. Results to date suggest that the information processing structure of the optimal control model has significant potential in this area but that much more work is required before this potential can be realized.

2.4 Models of Human Information Processing

The idea of dividing mental operations into component parts on the basis of latency measures has been employed off and on since the time of Donders (1869). However, it was Broadbent, (1958) in his seminal book on human information processing, who first popularized the idea that it is possible to model component processes in human information-processing systems as relatively independent entities. Building on Shannon's notions of information as a measurable quantity, Broadbent addressed the possibility of tracing the flow of information through the human sensory-motor system one stage at a time.

Parallel to these developments, a new field of mathematical psychology was evolving from purely statistical analyses of experiments and psychometric measurements to encompass the proposal of formal statistical models for learning (Bush and Mosteller, 1955), and statistical models for sensory processes (Tanner and Swets, 1954).

The work on models for learning slowly coalesced with information-processing theory to place emphasis on memory as a process rather than on learning per se. The resources of mathematical psychologists began to diffuse toward models for a variety of information-processing operations ranging from perception through memory and response.

During the last twenty years, a healthy interplay has developed between mathematical modellers, who began collecting data to evaluate their theories, and experimental psychologists, who sought quantitative models to explain their data.

A third thrust in the development of information processing models was provided by the advent of the computer as a tool for simulation. For those interested in psychology, this impact was felt in the formulation of computer simulation models for problem solving (Newell, Shaw and Simon, 1958) and other information-processing activities such as memory (Feigenbaum, 1961). Like the stage-modelling approach, these efforts reflect the convergence of several different philosophical, methodological, and technical threads.

The theoretical position of investigators in this area is that human problem solving can be understood and, therefore, modelled through characterization of three major dimensions: (1) the task environment, including the problem itself and the means (e.g., rules) available for solution; (2) the problem space employed by the subject to represent the task environment and an evolving solution; and (3) the "program" developed to achieve solution. The procedures used to characterize these determinants involve intensive prior analysis of the problem to be solved and equally intensive collection of subject protocols during solution. These techniques have included such problems as (1) choices among alternative moves in the game of chess; (2) cryptarithmetic problems, in which the problem solver must discover a code that substitutes a unique set of numeric characters for a given set of alphabetic characters; and (3) the development of proofs for problems in symbolic logic via formal application of appropriate rules.

The major goal of this research is to develop, from the protocols of subjects and the problem-behavior graphs to which these give rise during subsequent analysis, a computer program that captures and reproduces both process and solution. This goal

has been approximated in the formulation of the GPS (General Problem Solver) program, which appears to simulate quite well the performance of skilled subjects on a variety of logical problems. As a result, it represents, in the view of its authors, not only a valuable technical achievement but also a convincing and testable theory of human problem solving.

Today, the distinction among the empiricists, the mathematical modellers, and the computer modellers has blurred as the results of experiments tend to be expressed in models, as the complexity of the models calls for computer implementation, and as the computer modellers look beyond subjective protocols to experimental results as the basis for cognitive simulations.

2.4.1 Integrative Information Processing Models

In the course of our explorations of the problem of large-scale simulation models it has become clear that one of the major difficulties lies with the inability to generalize from predictions generated on the basis of one scenario to others involving changed levels for particular system variables such as number of operators, communications bandwidth available or number of objects being controlled. One solution to this problem may be to consider in advance the range of system conditions for which generality is sought, to carefully modularize the model so that those systems variables can be accommodated and then to proceed to more detailed design. Another solution, or perhaps one to be applied concurrently, is to build a very general model of human performance at a level of detail that makes very few assumptions about the specifics of system implementation and simply to build the system model around this general model of the human operator. It is to this end that we have reviewed the modelling work in human information processing

because that is clearly the goal of those involved in much human information processing research.

Perhaps the major conclusion we have reached in this regard is that the state-of-the-art does not provide the kind of integrative models necessary to produce such a representation of human performance that could be tailored to this class of application. There are two major deficiencies that need to be resolved before this kind of capability would be feasible: we need a better understanding of how to relate task requirements to elemental human performance capacities and limitations, and we need models of performance that are more comprehensive and less compartmentalized.

The problem of translating task demands into components that can be related to human performance models is essentially a taxonomic problem. We need a methodology for analyzing tasks that abstracts the essential human performance requirements in a form commensurate with available model structures. While this is possible on the basis of intuition and art if we stick to relatively simple tasks such as reading meters or moving switches or controls, more complicated tasks typical of integrative behavior do not yield easily to such analysis. Fleishman (Theologus and Fleishman, 1971) has proposed an approach to defining a task taxonomy that utilizes factor analysis to define human ability dimensions that could then be associated with tasks to be analyzed through the systematic collection of judges' ratings of the abilities important to each task. The method is ingenious, but would require extensive empirical research to derive the abilities and validate judgment scales for a broad enough sample of abilities to be useful for taxonomic purposes. Furthermore, the taxonomy that would result would not be compatible with predictive modelling of the

type under consideration in this report since it is ability-oriented rather than process-oriented. Teichner (1974) has defined a task taxonomy, building on Posner's informational approach, that is promising in the areas where there is much empirical data but of limited usefulness where there is not. Teichner postulates a series of information processing stages similar to those of most othe formulations and argues that some tasks bypass certain stages altogether while others cause certain stages to drop out as a function of practice. As a general approach, this seems quite productive, but applying the rules and deciding when to eliminate mediating stages remains, for all but the most elemental cases, an empirical discovery process. Thus, the articulation between task and information processing model remains a black art requiring extensive empirical validation.

With respect to integrative models, the literature is rife with block diagrams of relationships among the postulated stages of human information processing (e.g., Teichner, 1974; Norman and Rummelhart, 1970; Kahneman, 1973; Morton, 1970). Each author usually focuses on a detailed analysis of one or more subsets of stages that are then shown to model particular laboratory tasks effectively. The existing models emphasize simple verbal memory tasks, alphanumeric identification tasks, or the perceptual aspects of information processing, as will be discussed in more detail below. To our knowledge, no one has provided the specifications that would convert a comprehensive block diagram into a working model at this level of detail. While it might be possible to do so in a specific case, it is our opinion that the current state of understanding would not support building such a model that would be generally applicable at this intimate a level of detail.

As a methodology, the approach that seems most likely to achieve this goal is the modelling effort that is a part of the system development activities of the Naval Air Development Center under the direction of Cmdr. Robert Wherry, called the Human Operator Simulator (HOS) (see Abstract No. 21). HOS is to provide a computer simulation of human performance in a goal-oriented, task-processing environment (Strieb, 1975). It is used in conjunction with other programs such as the Human Operator Procedures (HOPROC), Human Operator Data Analyzer/Collator (HODAC), and the HOPROC Assembler/Loader (HAL) to provide the capability for full simulation.

Four high level functions are defined:

1. The decoder, which provides the translation from task requirements to human performance functions.
2. The multiplexor, which sets priorities for the order in which task requirements will be sequenced, based on an index of criticality.
3. The estimator, which is the sensing-remembering core of HOS and assigns time costs to the elemental human performance operations undertaken. It includes a memory module that is based on a strength model of memory performance having a short-term and long-term memory capability. The memory module is the major probabilistic element in HOS. The estimator also includes a model of the operator's physical and biomechanical operations such as eye movements, limb movements, etc.
4. The banker, which accumulates time costs and serves to interface HOS with the simulated hardware components of the system.

The decoder subsumes the taxonomic aspects of the modelling problem we have discussed above. While HOPROC provides a convenient means for entering such information, as far as we can discern, it introduces no advance in the technology of converting task requirements into performance functions. Using the NADC system, this activity still incorporates task analysis and the intuition of experts knowledgeable concerning human performance to derive the detailed specifications.

The essential performance modelling activity is subsumed in the estimator module. It is here that the elemental information and motion processing components are represented. Detailed descriptions of the models that have been used were not available at the time of this report. We believe that the ultimate success or failure of HOS will depend on the success with which these processes are represented, which is again the same issue that faces all of us interested in modelling. While HOS does provide an insightful framework and it appears that at this level some modelling has been undertaken, such as memory, eye movement times, and limb movement times, the predictions of information extraction time are still estimated on a case-by-case basis, where they cannot be assumed to be generically constant. HOS does advance the state-of-the-art in attempting to make these estimates at the level of individual processing activities, however, instead of at the larger task level typical of most network models.

An important feature of the modelling methodology represented by HOS is the cumulative incremental approach that is being taken to development and validation. Validation of HOS may be undertaken in successively larger units. We are aware of validation to predict meter reading and scanning performance. For this purpose,

the models required for this subset of tasks have been implemented and tested against real data. When these models have been tuned, the underlying human performance components will be available to predict larger units of behavior of which instrument reading is a subtask, and so forth. We believe that this is an admirably systematic, though time consuming way to proceed.

A second important contribution is the insightful and careful way in which the various components of human performance are being modularized and interrelated. As a result, many of the aspects of human performance fall out as a consequence of these logical relationships, thereby reducing the complexity of task modelling requirements.

As will be discussed in a later section, one cannot evaluate the success of a human performance modelling effort without understanding in considerable detail the goals to which it is addressed. Cmdr. Wherry argues that HOS is directed at system evaluation down to the level of detailed workplace layout and control and display design. As presently conceived, it is certainly capable of that level of detail. The question of interest here is whether it is also useful for evaluating more global aspects of complex system performance. On this point there is, as yet, no evidence. The modelling of more complex collections of tasks in HOS faces the same problems addressed in Section 2.2 on network models. It uses a bottom-up approach, and its success at complex levels depends on the modeller's ability to structure the tasks so that the main effects and interactions of interest are observable and measurable and are not buried in the implicit structure of the model. Supposing that global system performance measures were the variables of interest rather than workplace layout per se, one could make the argument that carrying the model to the level

of detail they propose is at best unnecessary and at worst impractical, in the sense that the effort required in time and expense to produce a system-level validated model is greater than the payoff that will result. It seems likely that models should be developed to provide answers to questions of interest and should only represent performance to the level of detail required for answering those questions. The quest for the ultimate model of human performance, a goal that one might read into the HOS efforts, may be unrealistic.

2.4.2 Information Processing Models of Limited Scope

Given these reservations about the development of a truly general model of human performance, let us drop back a level and ask what kinds of useful process models of more limited scope are available. We do not pretend to have conducted an exhaustive search of available models, and undoubtedly there are bits and pieces of models we have examined and chosen not to describe that might at a later time be shown to be useful. The information processing models for which we have provided abstracts in the Appendix are representative of the best we have been able to identify according to the evaluation criteria such as relevance, complexity, and usefulness in the topical areas identified in the introduction.

2.4.2.1 Information Measurement

Some of the first quantitative efforts to model human performance were derived from Shannon's information-theoretic concepts. It was these concepts that led to the idea that an operator's performance could be viewed as a limited-capacity, single-channel information processing system. Posner (1964) distinguished information-conserving tasks, in which there is a one-to-one mapping

between stimuli and required responses, from information-condensing tasks involving a many-to-one classification and from information creation tasks in which there is a one-to-many mapping of stimuli to responses.

In those information conserving situations in which the stimulus set and response set can be well-defined, either in discrete terms or in terms of continuous probability distributions, there is a very large data base and a sound basis for making predictions of the time required for processing. While this is possible in many practical cases, in many others the technical definition of the effective stimulus and the basis for quantifying it in informational terms is very difficult and often arbitrary (see Abstract No. 22).

As a measurement tool, information theory offered a unique way of combining data concerning speed of performance with that of accuracy into a single metric, and it was because of this feature that it gained considerable popularity in the late 1950's and early '60's. It is now recognized that the way in which this relative weighting of speed and accuracy were introduced represented only one of many possibilities and that it may not always be the most appropriate one. Further, the computation of rate of information transmission requires knowledge of the full matrix of data corresponding to the joint stimulus-response probabilities on a cell by cell basis. Where prediction of performance in new situations is the goal, as it is in the context of this report, the computation requires estimation of these probabilities in advance. It would be rare to have such advance information without having accumulated considerable experience with the system in actual operation. As a result, the predictive usefulness of the information metric is limited to cases in which error rates

can be assumed to be negligible, and the major concern is with the speed of information processing performance.

Finally, there are many factors that influence the rate of information processing that are not captured by the information metric. These include temporal uncertainty, stimulus-response compatibility, practice effects, stimulus intensity, and stimulus and response repetition effects. At the present state of our understanding, these factors must be dealt with as a catalog, assigning values to them on the basis of previous experimentation and judgment. We do not have formal models with which to integrate them into the overall information flow analysis.

Because of these limitations, research on human information processing has moved toward analysis of the various stages of processing that have been postulated for the human operator. It is to these classes of models that we turn next, beginning with those models that take a decision-theoretic view of performance.

2.4.2.2 Decision-Theory Based Models

In the decision-theoretic framework, the environment in which decisions are made is cast in terms of (1) a set of mutually exclusive and exhaustive but uncertain states of the world to which *a priori* probabilities are assigned, and (2) a set of alternative courses of action or decisions, also mutually exclusive and exhaustive, that may be taken. The observer accumulates evidence over time concerning the probabilities of the various states of the world and evaluates that evidence in the context of the values and costs associated with combinations of each state of the world and course of action.

When one is interested in predicting the probability of a perceptual response - target detection, recognition, and identification - Signal Detection Theory (Green & Swets, 1966) provides the most thoroughly formulated and generally applicable model. It is essentially a Bayesian model that revises the a priori probabilities of alternative states of the world based on human observation, and this information, together with collateral information concerning the values and costs associated with action, leads to a prediction of the choice of an alternative response (see Abstract No. 23). The elegance and power of the model derives from the fact that under conditions in which the environmental description of the signals and noise in which they are embedded can be expressed in quantitative terms, it is possible to predict the performance of an ideal statistical observer and to describe human observer performance in terms of deviations from this ideal. Even under conditions where this is not possible, it has been shown empirically that human performance frequently relates well to the underlying parameters of the model and that it may be generalized to pattern detection tasks and other recognition and identification tasks for which an ideal observer's performance would be incompletely specified.

Signal Detection Theory is usually applied to discrete-trial tasks, such as yes-no detection tasks or two-alternative forced-choice tasks. To apply it to a surveillance problem in which there is no inherent temporal structure requires making further assumptions about the frequency and duration of observations. Egan et al (1961) and Luce and Green (1970) have developed generalizations that can handle these circumstances. Lucas (1967) formulated a model of ideal performance for the case in which the time period of surveillance was limited, each observation incurred a cost, and the cumulative observation cost was limited.

Most applications of detection theory have been as a methodology for making quantitative measurements of sensory and perceptual capabilities rather than as a predictive model. However, there are several applications to visual inspection tasks, and the model has been shown to be applicable to vigilance behavior (Swets, 1977).

Closely related to the signal detection perspective is the work on latency models for human discrimination, identification, and response. Virtually all of these models take the view that evidence accumulation in the human information processing system takes time. Broadbent (1971), adopting the position put forward by Audley and Pike (1965), distinguishes among three alternative approaches: (1) accumulator theories with perfect memory, in which the choice of response is based on a race between evidence accumulation (in discrete units) in favor of each response, and in which the response corresponding to the first accumulator to reach a criterion accumulator count is the one chosen; (2) accumulator models with defective memory, in which the proportion of discrete "observations" in favor of a particular response must exceed a criterion, but only evidence through a particular "memory window" is accepted and older observations are dropped; and (3) random-walk models in which each new sample or observation causes a change in the relative likelihood of alternative responses and thereby causes a shift in confidence in favor of each response. Audley and Pike (1965) and Laming (1968) discuss the predictions of these models. Edwards (1965) presents the most completely formulated random-walk model, but the mathematics are only developed for the two-alternative case and it is not well validated.

Vickers (1970, see Abstract No. 24) presents an interesting variant of the random-walk model which incorporates the concept that sequential samples may vary in the magnitude of the evidence they contribute to a choice. With this formulation, he is able to show that accumulator-models and the standard random-walk model fall out as special cases. Here again, Vickers has dealt only with the two-alternative case and has validating evidence from only a limited class of laboratory tasks.

If one wished to model stimulus discrimination or identification in a practical setting, the ideas from these models would be useful, but it would require further model development and validation to ensure their applicability.

A slightly different approach to stimulus identification has been developed by several researchers who have been more interested in recognition and identification probability under tachistocopic conditions, that is, where the information is only briefly presented. The work of Rumelhart (1970), and Rumelhart and Siple (1974) is representative of the most completely formulated models for this purpose (see Abstract No. 25). There the identification process is viewed as one of selective feature extraction, with the number of features coded in the display time available leading to a prediction of the probability of pattern identification. The model has been shown to be successful for a remarkably broad range of laboratory conditions, but brief exposures are reflective of only a very small segment of the activity undertaken in any practically useful context.

2.4.2.3 Models for Short Term Memory

Perhaps the most active modelling area in psychology has been concerned with representing memory performance. Some theorists believe that the difference between short-term and long-term memory characteristics is only one of degree (Melton, 1963). However, most specialists now agree that retention and loss of information from short-term memory differs qualitatively from these processes in long-term memory. Some current theorists even propose a three-stage memory system including an intermediate or working-memory level in addition to the two usually defined (Broadbent, 1970).

Early formulations of memory models focussed on trying to settle the issue of whether an all-or-none process provided a better representation of the acquisition of information in memory as contrasted with an incremental or "strength" theory of memory (see Restle and Greeno, 1970, for a discussion of this issue and the models proposed).

A significant advance in the development of memory modelling came when Atkinson and Shiffrin (1968) formulated a model that for the first time distinguished explicitly the structural representation of memory stores from the task-dependent control strategies used for encoding information in these stores and interrelating them. (See Abstract No. 26.) This model also represents one of the most complete and, therefore, potentially useable models of memory performance.

Saunders, Smith, and Teichner (1974) have reviewed an extensive collection of other models of memory. We believe they would support the argument that none of the existing models are

formulated in sufficient detail and with sufficient generality to translate directly into a practical context. What they represent is a set of concepts that might ultimately be used in models for memory performance in a practical context. A further limitation is that they have dealt almost entirely with memory for verbal, alphanumeric symbols. The modelling literature for memory of pictorial material is virtually non-existant, although some interesting experimental studies are beginning to appear.

2.4.2.4 Motor Performance

Most of the interesting work modelling human motor performance has focused on closed-loop tracking control as discussed extensively in earlier sections of this review. The major exception to this view is the experimental and theoretical work related to what has been called Fitts' Law (see Abstract No. 27). This work provides a means for predicting movement time given the amplitude to be moved and the accuracy tolerance for the movement. The predictions have been validated for manual movements in various directions, for foot movements (Drury, 1975), and for head movements. Recently, Kvalseth (1975) has generalized the results to take account of cases in which the targets are temporarily obscured until the movement is actually in progress.

For simple movements, the model is robust and effective. As one proceeds to more complex movements, it may be necessary to augment this model with data derived from predetermined time systems used in industrial engineering, a procedure that operates with tabled values for movement time for a variety of well-defined, elemental manual movement activities (Karger and Bayha, 1966). Predetermined time systems have been developed in a manner analogous to the data-bank approach to task-centered predictions.

2.4.2.5 Intellectual Tasks and Mental Computation

As we move toward higher mental processes, models become more diffuse. Consider a question as seemingly straightforward as estimating the time to add two numbers together. If we restrict the domain to single digits whose sum does not exceed 18, then a counting model of the form suggested by Groen and Parkman (1972) is very suitable. The observer takes the larger of the two numbers and increments it serially at a rate of 20 msec per integer until the sum is reached. Seldom, however, would such a restricted domain be useful. Thomas (1963) has examined addition and multiplication and has shown that the log (a+b+c) where $a + b = c$ or $a \times b = c$ is a reasonable predictor of computation time. Again, however, the domain is rather limited and many assumptions must be made to extend this simple model even to encompass routine calculation requirements. Similar efforts have been directed at modelling the time for simple logical inferences but the results are specific to the circumstances investigated (set inclusion and exclusion) and are unlikely to generalize.

Considerable information has been accumulated concerning human ability to make simple intuitive statistical inferences (Peterson and Beach, 1967, (see Abstract No. 28). These data focus on the accuracy of such inferences rather than the time for such inferences. Kahneman and Tversky (1972) have described the sorts of heuristics that people use in making inferences, but thus far they can only be captured in qualitative terms (see Abstract No. 29). Both of these sources have more of the character of data appropriate to a data-store representation rather than that of a formal model, but they might be useful for the further development of quantitative models.

The one area in which considerable success has been achieved in modelling judgmental processes is that related to probabilistic inference under uncertainty. Again drawing from the decision-theoretic framework, the basic model is nothing more than Bayes' Theorem,and it provides a normative representation for how a decision maker should revise his estimates in the light of new data (see Abstract No. 30). However, some of the experimental work, such as Edwards, Phillips, Hays, and Goodman (1968), indicates the systematic ways in which human inference is degraded when data aggregation is required and could be used as a basis for a model. These ideas also have been extended to model the impact of unreliable data, a condition frequently encountered in the practical world (see Abstract No. 31). It should also be made clear what they do not do. They make no attempt to model the time required for judgment. They do not predict the specific likelihood ratios an operator would assign to each data item; such estimates would have to be anticipated and provided to the model.

Perhaps the most sophisticated attempt to model human reasoning processes are embodied in the General Problem Solver of Newell and Simon (1963). This work represented the peak of the era in which protocols of the activities of a human subject when solving a problem were described in a computer program. While some of the approaches taken in this program might provide useful concepts for the kind of modelling we are concerned with, the goals of this research were quite different. The authors' emphasis was on producing a program that would go through the same steps as a human problem solver. It placed less emphasis on predicting accurately the frequency with which the program would find a solution in comparison with human success probability. Further, the problems dealt with had a single unique solution, a circumstance not common in the inference contexts with which we are dealing.

Finally, the program was so complex that it made no attempt to mimic the time required for solution or to estimate solution times. Often the program took longer to solve the problem than real subjects did. Human problem solving behavior is so idiosyncratic that it seems unlikely that this approach would provide a productive strategy for predictive modelling of human performance.

2.4.3 Summary

We have examined as carefully and realistically as possible the prospects for developing models of human performance based on the human information processing modelling literature. We conclude; (1) that bits and pieces of relevant models exist, (2) that there is a rich experimental literature that can provide ideas and concepts that may be useful for building human operator models in specific task contexts, but (3) that no generally useful integrative models exist and that it is probably premature to try to develop one.

Advocating the use of the information processing data and modelling literature as a base presupposes that models will be developed beginning with the most elemental components of performance and building from there. If this approach is to be taken, the most promising strategy is one similar to that being pursued by the developers of HOS. We should pick a particular task or collection of tasks, analyze (on the basis of the best expertise available) the information processing components of each task, and assemble a model for each task in such a way that the functional components or modules of the model have generality even though they have been uniquely combined to represent only one task. These task models should then be validated against real human-operator data before proceeding to the next level of model aggregation across several tasks. Such an approach will require great

time and effort, especially early in its development when the most basic components are being formulated. This approach should provide valid and general representations for particular task aggregates, such as the RPV control problem, but would probably still require substantial new effort to apply to a new problem context.

The models derived by human information-processing specialists appear to have several limitations in their potential applicability to man-machine system modelling problems.

First, they tend to be compartmentalized by the very fact that most of them deal with particular stages of information processing rather than being integrative of human performance more generally. While this may make them useful for introduction as component models in a task-oriented simulation, one faces the same problems of independence and interaction at this component level that network formulations face at the task level (see Section 3.3 , below). Sternberg (1969) has offered the additive factors methodology as an experimental means of isolating those information-processing stages that operate independently, but in fact, the results are very dependent upon the experimental task or paradigm in which they are measured, and one has no assurance that the results will generalize without specific validation in each new context.

Second, human information processing models, for the most part, deal with the average performance of well-motivated highly practiced individuals under relatively ideal conditions. There are many hypotheses but few data and virtually no models in the information processing literature on how human performance capacities change under stress, under reduced motivation, before practice has

stabilized performance, when interacting in groups, or on the range or characteristics of individual differences. To the extent that these variables account for large proportions of the variance in practical situations, even the best models we can build will only be approximate.

2.5 Air Traffic Control

Our review has disclosed surprisingly few models relevant to human performance in air traffic control systems, though the literature on design of such systems and on estimates of the capacity of current and projected systems is very large. We have included in the Appendix summaries of two models - one concerned with perception of conflict and one with controller workload - that are relevant to the concerns of this report. (See Abstract Nos. 32 and 33, respectively.)

The model of conflict perception by Dunlay, et al (1973) deserves particular attention because of the author's attempts to address the discrepancy between prescribed behavior, as set forth in formal FAA regulations, and actual behavior, as determined from study of controller responses. An effort is made to accommodate the few empirical data available on controllers' perceptions of aircraft separation, using current display hardware. The model indicates that collision avoidance commands can be expected to be issued prematurely (as compared with the dictates of the regulations) and, hence, lead to greater controller workload than is customarily assumed in normative ATC models.

2.6 Models of Industrial Inspection

Perhaps because of their ubiquity and importance in industry, such tasks as allocation of resources, network capacity planning, quality control, etc., have been of increasing interest to operations researchers, applied mathematicians, and human factors experts. The result of this interest is a growing number of carefully defined and frequently elegant models.

A large percentage of the models we have reviewed in these areas are prescriptive in nature - that is, they are concerned with explication of optimal performance under given sets of constraints - rather than being descriptive of actual performance. This characteristic has, in our judgment, at least two important ramifications vis a vis application to modelling and simulation of human performance. One is that the models can probably be effectively employed only after the design of a given system is relatively complete and the goals of and constraints on its operation can be described in detailed form.

Secondly, and more importantly, since many of these models treat the human on a "black box" basis and do not attempt to accommodate empirical data relating actual performance strengths and limitations, they have limited value in identifying the most effective uses of humans in complex systems. Although one might, as the result of exploiting a particular model, gain some understanding of the best that can be hoped for from a given system, he is likely to be left uninformed as to how to realize that level in terms of real combinations of men and machines.

Notable exceptions to these generalizations occur in the area of inspector performance where, as McCornack (1961) noted, in his review of the literature, that the occurrence of systematic rather than random errors is of importance to the serious study of inspector accuracy. Efforts in this area have concentrated on classifying types of inspection activity - for example, examining each member of a set of similar items in an effort to identify one or more that do not conform vs. examining each member exhaustively for defects (see Fox and Haselgrave, 1969) - on empirical determination of what parameters of inspection tasks are critical in influencing accuracy, and on developing and applying modelling approaches that explicitly employ cost/payoff concepts.

A theoretical framework that has proved most useful in the modelling of inspection performance is that of signal detection theory (see Abstract No. 23). This framework, first applied to industrial inspection by Wallack & Adams (1969), was successfully used by Sheehan and Drury (1971); see also Drury, 1973) to study the commission of "false alarm" (classification of acceptable parts as rejectable) and "missed alarm" (classification of rejectable parts as acceptable) errors among a set of quality control inspectors in an experiment aimed at assessing the effects of illumination, visual acuity, and learning on performance.

Other existing models have been adapted to predict the results of inspector performance. Cochran, Purswell, and Hoag (1973) have developed four models based on Response Surface Methodology (see Clark and Williges, 1973, and Williges and Mills, 1973) for the purpose of predicting the relationships among rate

of change of visual angle, angular velocity, viewing time, illumination, and contrast in dynamic visual inspection tasks. Sheridan's (1969) model (see Abstract No. 34), based on Bayesian concepts, is another example of the adaptation of frameworks developed and used successfully in other contexts.

In addition to modelling techniques brought over from other areas of interest, models associated uniquely with the inspection and process control environments have begun to appear in the literature. One, formulated by Drury (see Abstract No. 35), accounts for an earlier finding by this author (Drury, 1973) that increases in inspection time are accompanied not only by an expected decrease in the acceptance of faulty items but also by an increase in the rejection of good items. A final example in this category is the model by Thomas (1973; see Abstract No. 36), which incorporates the linear relationship discovered by Hick between reaction time and number of decision alternatives, in a model designed to predict performance in a parts sorting task.

2.7 Miscellaneous Models

During our review, we encountered a number of models that, either because of their specific formulation or because of the situations to which they apply, were of only tangential interest to the context of this report. Some of these models are, however, well-developed and serve as good examples of modelling technology that might be tailored to the needs of the current project. We have included four such examples at the end of Appendix A.

The models discussed by Greening (1976; Abstract No. 37) have what is perhaps the greatest potential relevance of the set contained in the "Miscellaneous" category. In particular, they may be useful in the modelling of specific performance parameters associated with visual detection and recognition of targets during routine reconnaissance or during the weapons delivery phase of a mission. The remaining models are concerned with human performance in systems characterized by slow dynamic changes (Rouse, 1973, Abstract No. 38), with decision making in the context of alternatives that are dichotomous with respect to preference value (Pollay, 1970, Abstract No. 39), and with workload assessment under conditions in which attention must be shared between continuous and discontinuous tasks (Wingert, 1973, Abstract No. 40).

2.8 General Discussion and Summary

This review has examined a variety of models derived from a control-and decision-theoretic framework, the modelling literature in information processing, and the collection of models and data bank formulations originally derived from the reliability and network-simulation literature.

It should be clear from the preceding sections that manual control of vehicles has been the focus of a sustained national research program directed toward the development of human performance models of practical use for engineering design, and that efforts in this area have produced some well-defined and successful results. There is also a developmental thread in the area of network models and reliability models, but these models ave not been as broad-based or as successful thus far. The work on information processing models has been sustained, but its orientation has been toward advancing the theory of human performance rather than toward practical use, and this work has been fragmented and compartmentalized according to individual researcher's personal interests.

The most successful modelling efforts seem to have grown out of situations where formal models of the task environment are well developed, such as in feedback control tasks, detection tasks, and well-defined probabilistic decision-making tasks. Further, in these areas, the most successful of these models arise when the researcher can express formal criteria of optimal performance as reflected in the optimal control formulation of manual control or the ideal observer in target detection and recognition tasks. These observations emphasize the importance to successful modelling efforts of being able to express the

goals and success criteria used by human operators in formal quantitative terms. One difficulty for modelling behavior in more complex procedural tasks arises from the inherently multi-dimensional, multi-level, time-varying array of criteria and strategies that an operator applies in accomplishing these tasks.

It is interesting to note that the optimal control model and those information-processing models derived from a decision-theoretic framework are mutually compatible; this suggests the possibility of integrating and generalizing them to provide a single modelling framework that could be applied to vehicle control, supervisory monitoring, surveillance, signal identification, and decision making, all of which are tasks of major interest in military system design and evaluation.

As we view the modelling literature as a whole, the most productive work is in the area of signal detection, recognition, and identification, and in perceptual-motor control, where relatively tight feedback loops can be specified and human performance is relatively constrained. In the area of intellectual performance, modelling efforts have not produced practically useful results, either in areas where an explicit algorithm might be specified or with respect to general problem-solving performance, where a wide range of performance strategies are available. To represent these kinds of performance will require either structuring the problem so that results are not sensitive to differences in strategy or resorting to atheoretic representations derived from empirical measurements obtained in the specific task context.

A second area of difficulty concerns what Siegel and Wolf refer to as modulator functions (which change the over-all efficiency of performance) such as fatigue, motivation, practice, stress, or individual differences. While these can be dealt with superficially as simple modulators of performance efficiency, such a representation is largely atheoretic and fails to take account of the differential effects of these factors on particular aspects of performance and the interactions among processing activities that these differential effects imply.

One hopeful sign is that, for many purposes, models of performance may be robust with respect to such differences. Frequently the system and the environment impose constraints on human performance that restrict performance variability directly or limit the range of strategies or methods an individual may employ if mission goals are to be met. Frequently, models can be set up to take advantage of these constraints in predicting operator performance, and one finds empirically that operators adapt their over-all performance in accordance with the models. Where this kind of model is feasible, it may eliminate the need to deal directly with many sources of performance variation.

With the possible exception of SWM, we have identified no models that account satisfactorily for the compensatory interactions among participants in a team activity. A particularly difficult aspect of the team performance problem is representing the time and workload associated with interpersonal communication, since it is so highly idiosyncratic both within and between teams. Here again, however, the communication variability is typically reduced when the task demands build up, and this

feature of performance may reduce the importance of an adequate representation of interpersonal communication.

The models that we have reviewed exist at many different levels of specificity. A recurring theme has been the idea that systems models need to be built at a level of specificity appropriate to the goals of the modelling effort, and these need to be laid out carefully in advance. We believe that a network modelling language, such as SAINT, is adaptable to any desired level of specificity, ranging from the elemental implementation of an information-processing activity to the more global representation of task modules, where modules may be segmented at arbitrary levels. In Section 2.4, we observed that the careful segmentation or modularization of tasks and/or task components is of overriding importance, is made even more salient when we look over the diversity of modelling concepts that have been developed. The nature of the taxonomy cannot be considered without regard to the structure of the models themselves. The intrinsic feedback structure of control theory models requires segmentation of a different level than that appropriate for building tasks out of information-processing components. It is not clear whether the same hierarchy of structures that is appropriate to information-processing models can also be applied at the level of task-oriented models like that of Siegel and Wolf or the inspection models we have discussed. Finally, whether models from these diverse backgrounds can be integrated within a single task structure, as might be implemented in SAINT, remains a researchable issue.

In summary, we believe that integrative models of human performance compatible with the requirements for representing command and control performance do not exist at the present time. What is available is a collection of bits and pieces taken from a variety of frameworks that might be drawn together to build an eclectic model for a particular task situation of interest. The assembly of the pieces will require substantial effort in and of itself and is likely to require many assumptions about particular aspects of performance. If one is to have confidence in the product so generated, several iterative validation steps at different levels of abstraction will be required.

In the course of our review we have identified several issues relating to the overall perspective and implementation of systems models, and it is to these issues that we turn next.

3.0 ISSUES TO BE RESOLVED IN APPLYING HUMAN PERFORMANCE MODELS

3.1 Model Validation

The problem of validating a proposed man-machine model is a difficult one and, indeed, what is meant by model validation or verification is frequently debated. In one sense, a model is valid if it can be used to arrive at reasonable decisions; in such a case, accuracy may not be very important. However, with mathematical or simulation models, one usually looks for quantitative validation or tests of model accuracy. Even in this case, there are sometimes two views as to what is meant by verification. One view suggests that a model is verified when the model structure and parameters can be adjusted to provide an "adequate" match to experimental data. The second view holds that a model is validated only when the results of a new experiment are predicted by the model with sufficient accuracy. Here, we consider the first process to be one of model identification, and validation is taken to mean accurate prediction.*

For either view of model validation, it is necessary to compare experimental results with model predictions and to apply both engineering and formal statistical tests to determine whether or not the model should be considered "valid." To accomplish this, one must make judgments concerning (1) the definition of the "data," (2) the appropriate figures-of-merit for the engineering and statistical tests, (3) the specific statistical test to use, and

*The two views blur when one adjusts model structure and parameters to achieve a "match" to several experimental conditions simultaneously. If such a match is obtained with an invariant model, one can consider the model capable of prediction.

(4) the degree of discrepancy between experimental and model results that is considered acceptable.

For purposes of model validation, it is entirely proper to consider one of the experimental test conditions as "baseline" data for the purposes of adjusting human operator parameters of the model. In this way, one eliminates the possibility of rejecting the model because of differences between the subject pool available for the validation experiment and the subject population used in other studies. Instead, one can concentrate solely on testing the ability of the model to predict performance changes across the various simulation configurations of interest.*

One approach that is commonly used for statistical testing follows the "null hypothesis." Specifically, one adopts an initial hypothesis that discrepancies between experimental results and model predictions are the result of "experimental error" (that is, the result of inherent variability in the data and not of inaccuracies of the model). A formal statistical test is then performed to determine the probability that such discrepancies are due to experimental error. If this probability is sufficiently low (i.e., below some "confidence limit"), one concludes that model inaccuracies are statistically significant and the model is "rejected." Otherwise, the model is considered valid (more precisely, we fail to call it invalid).

*We assume that each test operator participates in each phase of the validation experiment. That is, each subject is used as his own "control"; otherwise, differences due to configurations are confounded with individual subject differences.

One of the inevitable problems with this type of testing is that the power of the test depends on the quality of the experimental data. For example, if the subject pool is not sufficiently large and/or if subject-to-subject variability is very great, it may be difficult to reject the model unless it is totally unreasonable. That is, we do not have a sensitive test to apply to competing models.

Therefore, the experimenter must conduct a very carefully controlled experiment so that the statistical test is reasonably discriminatory. Frequently, this is not a major problem for well-trained, well-motivated subjects performing well-defined, demanding tasks. For example, in laboratory experiments in manual control it is often the case that four test subjects provide sufficiently consistent data so that performance differences on the order of 15% can be considered to be significantly different at the 0.05 level of confidence. A somewhat larger subject pool may be desired for more complex C^3 tasks, since one expects subject variability to increase as the number of performance dimensions increases.

The choice of a confidence limit for rejection of the model depends on the tradeoff one makes between the probability of rejecting a model when differences are actually due to experimental error, and the probability of accepting the model when it is truly inaccurate. The lower the confidence limit, the greater must be the discrepancies to cause rejection. Typically, confidence limits are set in the range of 0.01 to 0.05.

It is also possible that in some unusual cases the statistical test may prove too strong. For example, if the subjects are extremely consistent in their performance, and if the subject pool is large, performance differences as low as 10% could be declared statistically significant, leading to rejection of the model. Accordingly, it is important that a valid engineering test be applied along with the formal statistical test.

Since there are a number of statistical tests one might use to validate a model, one cannot define a "best" procedure. One reasonable approach to statistical model validation is to apply t-tests to all relevant performance measures. In terms of a validation experiment

$$t = \frac{\bar{x}-\mu}{\left\{\frac{1}{N_s(N_s-1)} \sum_{i=1}^{N_s} [x_i-\mu]^2\right\}^{1/2}}$$

where N_s is the number of subjects, $\bar{x}_i$ is the ensemble average of the particular performance measure for the i-th subject, $\bar{x}$ is the average performance across all subjects, and μ is the model's prediction of the corresponding ensemble statistic, $\bar{x}(t_k)$.

Note that a separate test is performed for each ensemble variable of interest. Rejection or acceptance of the model along this performance dimension might then be determined by the variable that is least well predicted. That is, the model is rejected if the t-tests reveal at least one measure that cannot be attributed to experimental error.

The following figure-of-merit is often used as a possible basis for judgment of model validity:

$$E = \frac{1}{N_V} \sum_{x} \left| \frac{x-\mu}{\bar{x}} \right|^2$$

where E is the measure of prediction error, N_V is the number of variables of interest, and $\bar{x}$ is the data ensemble statistic for which μ is the corresponding model prediction. The summation is taken over all matched variables. Thus, E can be interpreted as the average squared difference in terms of percentage mismatch.

Alternatively, if sufficient data are available from different subjects to compute the inter-subject variability, then an appropriate engineering figure-of-merit for a given performance variable is a standard score, computed as

$$Z_{\bar{x}} = \frac{|\bar{x} - \mu|}{\sigma_{\bar{x}}}$$

where $\bar{x}$ and μ are defined as before and $\sigma_{\bar{x}}$ is the standard deviation (across subjects) of the performance measure $\bar{x}$. If $Z \leq 1$, then we can conclude that the model is about as good a predictor of subject performance as is another subject. This is probably all one can hope for. Of course, one has the option of averaging the Z's for different performance measures or of treating them independently.

As a last point, we consider the situation where the same performance variables are measured for a number of different experimental conditions. One can consider the deviations of these measurements from model predictions in a manner similar to the above treatments. Alternatively, one could use the correlation statistic as a quantitative means of examining the correspondence between model predictions and empirically-derived results. For

each performance measure, one would compute the correlation coefficient between observed performance and predicted performance across a set of experimental conditions that give rise to a suitable range of variation in the measure under study. To provide the strongest test of the model, one would want to either fix the parameters of the human operator portions of the model and vary only those parameters that are associated with the experimental conditions under study or at least to select human operator parameters according to rules defined in advance. The more human operator parameters that are adjusted specifically to match the data, the easier it should be to obtain high correlations.

For models of complexity sufficient to represent full-scale man-machine system performance, the problems of validation go well beyond the selection of the proper goodness-of-fit statistics. If the model is to be useful as a design tool, it must be validated prior to, or at least concurrently with, the development of a full-scale simulation. Under these conditions, system performance data against which to validate the model are not likely to be available, and one must seek less rigorous means of examining the validity of the model. In the course of building the model, the investigator will inevitably generate qualitative hypotheses about how behavior will change with changes in certain critical system parameters. At a minimum, one should be able to predict the direction of changes in system performance measures if not their magnitude. Thus one simple step toward validation would involve testing the model against such system parameter changes to ensure that it responds in accordance with these prior expectations. It should also be possible to define limiting conditions and to confirm that the model behaves properly in these cases.

There is another sense in which even a full scale empirical test would have limited usefulness. If the system model is complex and has many levels of detailed representation, collecting data concerning changes in system performance as a function of parameter changes will tell us only whether the model is successful (an unlikely outcome on the first try), but not what to do to improve it. The empirical validation must either proceed in steps from specific to general, as suggested in Section 2.4, or else provide data concerning performance at all levels in the hierarchical structure of the model and in terms of the identifiable model parameters at each level.

Finally, there is a logical paradox that arises when attempting to validate any predictive model, the purpose of which is to describe performance in some new situation; if we experimentally examine conditions sufficiently close to that new situation, we have eliminated the need for prediction, since we would then have available experimental data of direct usefulness. Perhaps the goal of validation should be to span the range of conditions of application so that prediction becomes a task of <u>interpolation</u> in a multi-dimensional parameter space whose boundaries have been explored, rather than an <u>extrapolation</u> to a wholly new domain of activities.

In summary, the problem of validating large scale systems models is not one for which there are always textbook answers. It is an issue that must be addressed for each new model that is created, and further effort is required to develop guidelines for how to proceed in particular cases.

3.2 Selecting the Level at Which to Formulate Man-Machine Performance Models

There are several issues that interact in the attempt to select the appropriate modelling approach to solve a given system evaluation problem.

The overall goal should be to develop a model that has sufficient generality to be able to predict system performance over a range of parameter settings for classes of parameters or attributes for which the designers have some leeway in design and for which they are uncertain regarding suitable values. It is very difficult, if not impossible, to develop a model that will serve as a useful design tool for the study of circumstances that were not anticipated at the time of its conception without some modification or adjustment along the way.

Thus a modelling effort should begin with a clear definition of the purposes for which it will be used and the conditions under which validity is desired. A critical part of this definition is the level of detail at which prediction is desired. If the primary objective of a model of an RPV Control System is to predict the effects on weapons delivery performance of changes in communications bandwidth between the control station and the vehicle there is no need to represent human performance at the level of detail that would be sensitive to changes in workplace layout. A model that is to evaluate operator workload, however, may require a representation of the workload associated with the operation of a light pen that can be invoked each time a light pen is ued, but it need not represent the perceptual-motor details of light pen operation. On the other hand, if a design

change will impact on the way a light pen is used as opposed to the number of times it is used, then a more detailed representation may be required. We know of no general procedures that will replace thoughtful analysis for arriving at a judgment of the suitable level at which to model.

The question of level of detail interacts with another issue which remains to be resolved: when modelling a complex man-machine system, does one proceed to develop a block diagram at the most general level and then break it down into more detailed levels as necessary or does one start at the most fundamental level of behavior and systematically integrate these components to build the model? These two approaches are sometimes referred to as the top-down vs. bottom-up approaches or the analytic vs. synthetic. Although clear distinctions cannot be made, the optimal control model has more of the flavor of the top-down approach whereas the HOS simulations are representative of a bottom-up approach.

It might be argued that a detailed analysis of the elemental components of performance, carefully represented and validated, could then be built up systematically into a more integrated model, as was argued in the discussion of information processing models. If all behavior were purely additive in time and multiplicative in accuracy, this idea might be practical. Unfortunately this is not the case. Behavioral scientists have always been cursed by statistical interactions among variables of interest, and such interactions usually must be modelled explicitly at each new level of integration that is undertaken. Because of human strategy variations and other vagaries of human performance, it simply is not possible to blindly assume additive and multiplicative combination rules without some a priori theoretical reason or an empirical validation of the result. This issue is addressed separately

in Section 3.3. It is this difficulty with strategy variations that limits the apparent generality of a bottom-up model. As soon as one examines new situations, new strategies emerge that require adjustment or reformulation of the model.

Perhaps the most difficult challenge of the bottom-up approach is the production of a suitable general taxonomy for deriving the components of human performance out of which the tasks are to be reconstructed. Unfortunately, we still do not have suitable procedures for developing such a taxonomy on any basis other than informed intuition. This is an area in which much work is needed, but it is not clear whether a method can be derived or whether the difficulty is simply reflective of our lack of theoretical understanding of the components of human performance.

If the bottom-up approach is thus restricted in generality, what of the top-down alternative? In this case one begins with a statement of goals or objective functions which the system is to minimize, a description of the design parameters, and the range of values over which generality is sought. The designer moves down the hierarchy of goals and subgoals only to the level at which the design parameters can be expected to have an influence. If such models are operating in the domain of well-developed theory, this approach may work, but we believe such theory is rather limited in scope and applicability at the present time. Further, when the performance goals are qualitative in nature the models may not be appropriate at all. Insofar as the models are prescriptive they are essentially incapable of predicting idiosyncratic behavior, as might be induced by such qualitative goals (although a range of acceptable behaviors might be predicted).

Finally, because top-down models do not get to certain elemental levels of modelling, they are frequently unable to predict the effects of changes that are the traditional concern of human factors specialists (e.g., display readability, control-display compatability, etc.)

Top down models reduce the taxonomic problem, but in its place a critical need arises for well-defined dimensions or parameters of system performance from which to formulate performance indices. It may be just as difficult to derive these for human performance characteristics as it is to derive the taxonomy.

We do not see this top-down vs. bottom-up issue as one having a general solution, but rather as one for which accumulated experience and validating data would help to guide the model builders working in this area.

3.3 Approaches to Handling Interactions Among Tasks

It was observed in connection with network approaches to prediction that there are few principles to guide the selection of composite reliability and time distributions. This matter is of concern where tasks or behaviors occurring in a procedure sequence are not independent and, hence, the standard product and summation rules of reliability theory are inappropriate. In such circumstances, application of these rules may result in predictive estimates that deviate critically from those derived from actual observation, as demonstrated by Buckner and McGrath (1961; see also Swain, 1964).

The significance of the problem has been noted by a number of investigators (Lamb, 1970; Meister, 1971; Swain, 1967; Mills and Hatfield, 1974) and with respect to the product rule, there appears to be general agreement over four possible solutions:

1) Find means for identifying and qualifying task dependencies and then reformulate the product rule to include a dependency term.
2) Limit use of the rule to situations where dependencies do not exist.
3) Limit use of the rule to situations in which large errors in predictive accuracy are tolerable.
4) Identify a level of performance modelling at which dependencies do not exist and the product rule can be used.

Of these, perhaps only (1) and (4) are solutions in the real sense of the term, since to limit modelling activity to problems where dependencies do not exist or to those where accuracy is not critical may be to constrain severely the set of man/machine systems one can hope to study. As a result, efforts to solve the dependency problem usually emphasize either the collection of molecular data under known conditions of task interaction, for the purpose of ascertaining appropriate corrections to the product rule (as in Mills, 1970; Mills and Hatfield, 1974), or the redefinition of molecular elements into task elements of higher order (as in Swain, 1970).

Both of these approaches have their strengths and weaknesses. The Mills and Hatfield proposal, which involves a program of parametric studies that examines in great detail the sequential dependencies among ongoing series of task elements and seeks to find distributions that adequately characterize combinations of these dependencies, holds the promise of yielding results that enable very precise prediction of performance. At the same time, the results may lack generality unless the program of research is exhaustive and includes the study of task performance in its full dimensionality - that is, with the accompaniment of motivational factors, stress, learning, etc., normally encountered under actual field conditions. Swain's approach, on the other hand, appears to promise generality but requires a degree of understanding of the detailed tasks and procedures to be pursued in a given system and a level of "artistry" in applying past experience with similar tasks and procedures that may preclude successful exploitation of the modelling technique by less-than-totally-informed analysts.

Our review has disclosed little that would suggest that, at this point in the development of modelling technology, one of these approaches is to be preferred over the other. The choice among possible combinatorial rules and distributions is so intimately tied to questions concerning the level of predictive accuracy desired, the adequacy of available process/task taxonomies, and intrinsic characteristics of each particular performance sequence to be modelled, that the only reasonable proposal is to continue the development of both approaches and to apply each where it seems to work most satisfactorily.

3.4 Underdetermination of Multi-Parameter Models

In a typical well-formulated mathematical model of a well-defined process, a goal of the modeller is to represent behavior with as few free parameters as possible. Ideally, there should be fewer such parameters than there are dependent variables to be predicted. This principle has been expressed by the epigram, "If we model the muscle system with 69 parameters it is possible to make a muscle spindle sing, 'God Bless America.'"

When we move to models of complex, highly-interactive man-machine systems, the issue is much harder to pin down. Model parameters generally fall into four classes: (1) Parameters that are defined by the initial conditions under study, such as hardware variables for which specification forms a part of the problem statement. (2) Parameters that form an integral part of the model, but that may be assumed to be invariant over the range of conditions to be studied. Their values may be estimated by validating experiments, by theory, or by assumption, but they are not free to vary from run to run. (3) Parameters that may vary from condition to condition, but for which theory or experiment defines the rules of variation contingent on the context in which they are to be assigned. (4) Free parameters that are set at the time the model is exercised in order to produce the best fit to data of that particular condition.

If a model is to be predictive, values should be assignable *a priori* to all these kinds of parameters. However, models might produce useful conclusions even when certain of the parameters are poorly defined, as long as predictions are derived on the basis of a range of reasonable parameter values. These results would then be used like sensitivity analyses to set best and worst case predictions.

An interesting issue for which we know of no guiding principles is just how constrained such a model must be in order to make useful generalizations. Is there a reasonable ratio of unconstrained parameters to dependent variables that leads to useful models? How much uncertainty in parameter settings can be tolerated before the predictions lose credibility? It seems likely that some of these questions depend on the particular model or application domain, but we believe some general statements might be made as we accumulate experience with alternative model forms.

Simple models have the potential for exploiting statistical procedures for identifying parameter values to maximize the goodness of fit to a data set; however, simulation models of the scope considered here can become a "fiddler's paradise" because of the complex interactions involved. Is it possible to devise systematic approaches to estimate parameters that do not have full statistical rigor, or even the rigor of efficient hill-climbing algorithms, but yet provide some bounds on the time, effort, and confidence in the values obtained? We regard this also as a researchable issue.

3.5 Potential for Systematic Model Construction

At the current state of technology, each new task situation requires beginning with first principles to derive a model to represent it. Until this situation changes, the time and effort required for producing such a model may exceed the benefits to be derived from it. Until we can systematize the process of model construction, as proposed by the HOS approach, or formulate a general system or simulation theory that provides a basis for structuring new models, progress will be slow and unrewarding. Research is required to identify the general

structure of system models and to provide a language for discussing them in generic terms before the community of scientists can take a more goal-oriented approach to model development.

4.0 RECOMMENDATIONS

In this report we have considered a large number of human performance models and modelling approaches. Further, we have considered a family of issues that surround attempts to employ these models in the predication of task performance. In this section we summarize our recommendations for future work concerning modelling of multi-man command and control systems. These recommendations include development of methodological tools and concepts as well as specific extensions of existing modelling technology.

4.1 Develop a Test-Bed Facility in Which to Evaluate and Validate Models on a Comparative Basis

While many models have been developed from a variety of perspectives focusing on a variety of issues, we know of only a few cases in which specific model validation studies have been undertaken, and we know of no cases in the system context in which the same operations have been modelled from different approaches in order to compare the relative effectiveness of the approaches employed. It is with these goals in mind that we propose development of a generalized test-bed facility.

At this point, we have not identified in detail the critical features of such a computer-based test-bed, but we will discuss some design considerations. (1) The test bed should have the capability of supporting a variety of modelling approaches, including what we have referred to as top-down and bottom-up models. (2) It should make possible multi-person "live" simulations as well as models of the equivalent manned systems so that validation studies can be conducted. (3) In order to facilitate implementation, it should incorporate a common modelling language and

operating system in which both models and live simulations can be rapidly constructed and debugged.

We envision four modular components to the operating system: (1) a module for implementing the hardware components of the system exclusive of the specific displays and controls with which human operators would interact; (2) a module that provides the software linkages to the specific display and control hardware to be used in manned simulations; (3) a module in which to implement models of human operator performance that takes as its input the same information that would be input to the display hardware and provides outputs identical to those produced from the operators' controls (this module would thus permit the same system hardware simulation to be used interchangeably with both manned simulations and models of the human operators' performance)[1]; and (4) an automated performance measurement module that would make it easy to program the data collection and analysis activities associated with either mode of operation so that the same measures could be collected in both cases when comparisons are to be made. In addition to the direct comparability that such an implementation would afford, a further advantage would be the ability to replace human performances of particular tasks or perhaps of one operator of a team with real time models of that performance, maintaining the rest of the simulation intact.

One of the primary uses we envision for the proposed test-bed facility is the implementation and comparison of various top-down and bottom-up model approaches for a wide range of task situations. A major goal of the research to be conducted would be to document and study the points at which these distinct modelling approaches run into difficulty; a subgoal would be to produce case studies of methods for integrating models that grow out of alternative approaches.

[1] In order for both simulations and Monte Carlo models to be practical utilizing the same hardware simulation module, it would be desirable to disable the real-time display-driving software and to run the hardware module in non-real time.

For those modelling approaches that involve synthesizing higher-level models from elemental components, we would urge that a process of incremental model-building and incremental validation be studied. This process would involve validating each of the model subcomponents as they are developed. We believe that this approach would yield several benefits in terms of both accuracy and simplicity. By conducting sensitivity analyses for various subcomponent parameters, one can probably reduce the number of variables that must be carried along into the higher-level components. If this turns out to be the case for a wide enough variety of contexts, the "fiddler's paradise" problem noted in the previous section may be substantially reduced.

A second benefit of this approach is that it provides a vehicle for systematic study of combinatorial problems associated with the aggregation of non-independent sub-tasks. As suggested at a number of points in this report, we consider these problems to be among the most critical facing successful exploitation of the bottom-up approach. Further, we believe that if these problems cannot be solved in principle for a significant number of system-relevant subtasks and their interrelationships, this approach may eventually give way to approaches that do not require the synthesis of subtask models.

In our discussion of approaches to handling the interaction problem (Section 3.2), we noted two possible directions: (1) a program of parametric studies attempting to uncover distributions appropriate to the combination of subtasks of known dependence, and (2) the redirection of a given modelling effort to task levels where interaction is not a problem, so that standard rules and distributions can be employed. We also noted that a clear preference between these alternatives cannot be established on the basis of our

review of the human performance modelling literature. The incremental modelling approach, pursued over a set of well-defined case studies, should enable one to gain a very clear understanding of the families of combinatorial rules and distributions that would be of general utility in bottom-up modelling efforts. Further, it should provide one with some feeling for the loss of accuracy occasioned by the use of simpler combination rules and distributions drawn from standard reliability theory.

Finally, the incremental approach offers a context in which to develop criteria and methods for fabrication of process/task taxonomies. One might hope to learn from these efforts whether it is possible to develop a single strategy for partitioning tasks and to employ successfully a single generalized taxonomy, or whether it is necessary to evolve a set of strategies and taxonomic procedures.

We cannot overemphasize our feeling that the generation of a testbed facility at this time is critical to the continued development of human performance modelling. We believe that it provides the only mechanism by which important issues can be clarified and needs for further data, theory and research identified.

4.2 Research Recommendations Having Applicability Independent of Modelling Approach or Topical Area

4.2.1 Component Model Aggregation

The use of models embedded in a larger network of tasks, whether they be models of tasks or models of component information processing activities, produce predictions of system performance through aggregation of the component activities. Research is needed to consider the impact of propagation of errors in component or task level models

on the overall accuracy of system level predictions as a function of the number of steps of aggregation. Swain, (1969) has provided a starting point for such a study in the area of reliability aggregation, but we believe the question could be addressed more generally to include propagation of other kinds of modelling error in both time and accuracy.

The modelling errors may be caused by biases in component predictions or by incorrect assumptions concerning the form or variability of the assumed component performance distributions. Some kinds of distributions may be more robust with respect to aggregation than others. Mills and Hatfield (1974) provide a starting point for this exploration.

A further aspect of error propagation relates to the structural interrelations among the component models and whether or not the interactions between components are considered explicitly, implicitly or neglected altogether. A systematic study of these influences would contribute generically to future model development.

4.2.2 Validation Research

As suggested in Section 3.1, the validation of models of the complexity and scope required to represent command and control systems remains a largely uncharted area. What are needed are methodological studies at more than one level of complexity in which model predictions and systematically degraded model predictions can be produced to examine the usefulness of several validation concepts.

4.2.3 Empirical Evaluation of the Problem of Parameter Underdetermination

Research is needed to examine the impact of trade-offs between the number of parameters that must be estimated from data in live simulation runs and the number of system performance measures to be predicted from the model of human performance in the simulation.

4.3 Research Recommendations Concerning Topical Areas for Model Development

We have identified two areas in the continued evaluation of man/machine systems where we believe there will shortly be a crucial need for modelling techniques. These are discussed briefly below.

4.3.1 Supervisory Control and Monitoring

Many of the activities that are performed by men in command and control systems may be thought of as supervisory monitoring and controlling activities. While applicable integrative models for these processes are not currently available, many of the concepts and principles have been introduced either in the manual control literature or in the information processing literature.

A recent NATO Conference (Sheridan, 1976) focussed attention on this topic, and we believe that substantial progress can be made. The requisite submodels include signal detection, visual scanning, decision, workload assessment, and control execution, albeit on a longer time scale than heretofore examined. We have described candidate models for all of these processes, which could be integrated into a common decision-theoretic framework that would provide a coherent description of both system variables and human performance variables.

4.3.2 Team Performance

Though conceptualizations of group performance abound in the literatures of social psychology and management, few are sufficiently quantitative in nature to be applicable to descriptive modelling of team performance in complex man-machine systems. For purposes of simulation, the Siegel & Wolf approach mentioned briefly in this report represents perhaps the most comprehensive attack to date on multidimensional aspects of crew performance.

We believe that the successful development of team performance modelling will be aided by inputs from at least two sources: (1) a reasonably comprehensive empirical data base, and (2) a conceptualization of team activity that clearly distinguishes those components of the total variance in performance that are associated uniquely with the interaction of team members from those that are associated with members acting individually; that is, a conceptualization of the general form

$$\sigma_T = \sum_{n=1}^{N} \sigma_n + \sigma_t + \sigma_e$$

where σ_T = the total variance in performance
σ_n = the variance in total team performance due to member n
σ_t = the variance in total team performance due to interaction among members
σ_e = the error variance

Efforts to meet these needs should proceed concurrently and with the pragmatic goal of determining what characteristics of the performance of n-person teams can successfully be modelled by combining the outputs of existing models of individual behavior and what characteristics require specific group formulation.

We believe that the generation of group performance data and models might usefully begin in the areas of decision making and system monitoring, for several reasons:

(1) Almost all systems that are not completely automatic in operation and self-monitoring rely significantly on humans to detect out-of-tolerance conditions and to formulate and/or choose among alternatives for correcting those conditions. Many systems, in fact, are designed around the notion that humans are better able to perform such functions than are hardware or software mechanisms, particularly where redundant procedures for monitoring and decision making can be devised. To our knowledge, there are few data that either affirm or disaffirm that notion, and no descriptive model of group performance is available on which to base a prediction that can be compared with the predictions of models of hardware and software performance.

(2) Prescriptive models of group decision-making and monitoring already exist in the literatures of decision theory and operations research and can be employed as yardsticks against which to measure actual performance or that predicted by descriptive models. With the ability to develop such a comparison, one then has a mechanism for assessing the costs and payoffs associated with system redesign.

(3) That such factors as motivation, training, stress, workload, etc. have significant differential effects on the decision making performance of individuals is

well recognized, and it is to be expected that these factors operate in like fashion on collections of individuals. What cannot be predicted reliably at this point is the cumulative impact of such factors on group productivity and output. Further, we cannot anticipate in more than a qualitative way the influence of variables such as leadership, which are uniquely associated with group enterprise. Empirical data and the definition and formulation of predictive models are critical here if future operating teams and crews are to be assembled on other than an ad hoc basis.

4.4 Advancing The State-Of-The-Art with Respect to Specific Modelling Approaches

In the course of our review we have also identified specific *recommendations* that would further the development of particular models and make them more applicable to command and control problems. These recommendations are listed below.

4.4.1 Data Bank and Network Approaches

1. Evaluate the usefulness of reliability measures derived as point estimates of the probability of successful task completion, as compared to error-rate specifications such as expected mean-time-between errors or failures.

2. Develop a systematic taxonomy for defining tasks and segmenting them into units for which performance can be measured. Examine the impact on modelling success of representations at different levels of task aggregation.

4.4.2 Control Theoretic Models

1. Extend optimal control attention-sharing model to apply to non-stationary situations.

2. Develop methods for applying manual control models to multi-operator situations.

3. Develop more formal procedures for specifying control and decision cost-functionals for the optimal control model.

4. Develop rigorous schemes for estimating statistical confidence limits on the parameters of the optimal control model. Evaluate and extend formal parameter optimization procedures.

5. Extend models that account for the effects of work-induced stress on control and monitoring behavior.

6. Extend methods for integrating the representations of attentional, display-scanning, and control workload.

4.4.3 Human Information Processing Models

1. Conduct research to yield a taxonomy that relates elemental human information processes to human performance requirements in realistic systems contexts. Develop theory that permits task requirements to be expressed in terms of component information processing activities.

2. Examine the linkages between detection theory and control-theory-based decision models to articulate a single model applicable to dynamic decision-making activities.

3. Adapt existing human information processing stage models and apply them to the representation of practical tasks. Evaluate their usefulness for this purpose. Candidate models include:
   a. Evidence accumulation models.
   b. Perceptual recognition models.
   c. Models of verbal memory.
   d. Models for algorithmic problem solving such as mental arithmetic.

Index to Appendix

ABSTRACTS

Data Bank Formulations

NOTE: Though not "models" in the sense in which that term is used elsewhere in this report, human performance data banks are of demonstrated utility to larger efforts aimed at estimating the reliability of various man-machine configurations. Because they continue to be of value to designers and analysts and because, as a group, they make explicit a large set of operator behaviors, performance criteria, and error types of interest to modelling activities, two prominent examples of the data bank approach are included here.

References:

Payne, D. & Altman, J. W. An index of electronic equipment operability: Report of development. American Institute for Research, January 1962.

Rigby, L. V. The Sandia human error rate bank (SHERB) Human Factors Society Symposium - Los Angeles Chapter, 1967, (5-1) - (5-13).

Swain, Alan D. Shortcuts in human reliability analysis. Paper presented at NATO Advanced Study Institute on Generic Techniques in Systems Reliability Accessment, U. of Liverpool, England, July 16-27,, 1973.

Description

Data bank formulations provide a mechanism for prediction of performance in prospective systems through the use of tabular arrays containing historical laboratory and field data on human reliability. Though the formulations differ in specific details, all attempt to capture and represent performance in subtask units that can be identified in a variety of system contexts (e.g., component inspection, manual operation of switches, etc.). Most data banks are concerned with at least two aspects of subtask performance: (1) the reliability (or alternatively, unreliability) with which a given subtask is performed and (2) the time taken (on the average) to complete the task. A final generalization is that most formulations contain, in addition to tabulated data, a rationale for the combination of subtask execution times and/or reliabilities in such a way that larger performance aggregates can be analyzed. This set of characteristics makes the data banks very similar in concept to those employed in prediction of hardware

reliability. The similarity is of benefit in an analysis of systems design concepts where there is a desire to evaluate human performance in the same terms as equipment performance (e.g., failure rate per n operations).

Data Store (AIR)

The first of the data banks to be formulated and one that continues to be of utility to performance prediction is that developed by AIR (American Institutes for Research) in 1962. Called "Data Store," this system breaks each step or action involved in a procedure into three factors: (1) nature of the stimulus or input to which the operator must respond (e.g., flashing of a light, movement of a pointer on a scale, etc.); (2) nature of the perceptual or cognitive mediating processes required (e.g., decision making, short- or long-term memory. etc.), and (3), nature of the output required (e.g., movement of a control, speaking, etc.). Associated with each of these is a measure of the time taken to respond in the required way that has been derived on the basis of laboratory studies and available field data, along with an estimate of the rate at which errors occur in that activity. Where an activity can be parameterized and the precision of the estimate depends on the values of the parameters, "Data Store" provides a "base-time" for the basic activity and a "time added" for each of the critical sets of values on each parameter. When using the system, the designer first chooses the activity that is most similar to that he is attempting to predict and then adds to its "base time" all "time added" increments that apply. Aggregate reliabilities are handled similarly, except that reliability figures for the critical values on each parameter are multiplied rather than summed.

SHERB (Sandia)

Since "Data Store," many human performance data banks have been developed. Here, we shall mention only one other, the SHERB (Sandia Human Error Rate Bank), developed and used in connection with THERP (Technique for Human Error Rate Prediction) by Rigby, Swain and Rook.

SHERB is a compilation primarily of four major categories of error: (1) assembly errors, committed in the production of electronic components and equipment; (2) installation errors, committed during installation and/or integration of smaller

units into larger units (e.g., sub-assemblies into assemblies) of equipment; (3) operator errors, committed in the course of operating, transporting or handling equipment; and (4) errors committed in the course of servicing and maintaining equipment.

Each of these categories is further divided into subcategories that identify more or less exactly the task of interest, the type of error of interest and the criterion for assigning the performance to a given error category. Against these are posted the major statistical descriptors, mean, standard deviation (or other available measure of dispersion), range of error rates, and (where identifiable) shape of distribution, available from laboratory and field observations relevant to that performance-task-error combination.

To aid in sensitive exploitation of the data base, additional information concerning the subject/operator population, work environment, climate, etc. in which tabled data were accumulated is abstracted and maintained as part of the data base. Certain dimensions of this supporting data, thought to be critical to interpretation, are rated on a seven point scale corresponding to the range from -3σ to $+3\sigma$ of the normal distribution in an effort further to assess their influence in producing observed results. Finally, evaluations of reliability, validity, generalizability and observer credibility, and of any other qualifying information relevant to task, nature of error, situation, etc. are made.

Like Data Store, SHERB can be used to supply data required for network analysis or for assessment of single task performance. In principle, one simply seeks a match between the activity (or activities) to be studied and the keyword and descriptive structure of the data bank, acquires data posted against relevant task-performance parameters, uses the supporting information to achieve a fine-grain correspondence between situational conditions and to qualify the tabled value, and makes his estimate.

Input Parameters

Data bank formulations do not have "Input Parameters" in the normal sense of the term. What is required is a fairly precise idea of the characteristics of the prospective tasks and subtasks so that appropriate portions of the data base may be accessed.

Model Outputs

Model outputs are typically (1) the time required to each member of a set of specifiable subtasks and the aggregate task time; (2) the probability associated with success (or failure) in the execution of each subtask and the success (or failure) of the task aggregate.

Model Validation

Since the data bases of the various formulations are constituted from existing laboratory and field data, they are inherently valid, at least over the range of conditions represented in those studies. Their validity when applied to new situations is limited by the objective similarity of the prospective tasks and conditions to those under which the data were collected in the first place, as well as by the sensitivity of the user in classifying the characteristics of a given task for purposes of relevant data base information.

Comments

The idea of tabulating results from laboratory and field observations and of using these to estimate performance in a system yet to be designed is clearly attractive, since it offers the possibilities of great economy (over, say, simulation methods) and a standardized methodology for reliability analysis. Associated with all data banks formulated to date, however, are a number of major problems. One is the fact that many of the data contained within them represent empirical outcomes of laboratory studies whose generality may be suspect. A second is the fact that much artistry may be required to match tabled descriptions of the tasks against what may be only vaguely conceived future tasks. A third is the likelihood that major doubt concerning the operation of a new system will center about tasks for which there is no directly relevant antecedent experience, hence no counterpart in the data base with which to estimate performance.

Despite these problems, the data banks may often provide one with a ready means for assigning an order of merit to each member of a set of design alternatives, and aid in pinpointing areas of the design that will require careful consideration, empirical study, etc. In this mode of use, they are probably indispensable.

Siegel & Wolf Model of Performance Under Stress

References

Siegel, A. I. and Wolf, J. Jay. Man-Machine Simulation Models. Wiley and Sons, 1969, pp. 17-18.

Description

This model, originally derived for use by the authors in their simulation technique, is based on the Cannon-Bard theory of emotion which views mild stress as producing a facilitative effect on performance and extreme stress as producing an inhibitory effect.

The primary assumption of the model is that "the 'certainty' in the operator's 'mind' that there is insufficient time remaining to complete all [the] essential subtasks when performing at normal speed and efficiency will cause a state of stress on the operator" (p.17). Two parameters are critical: (1) the average time required to complete the subtasks in question as compared to the time available to complete them, and (2) the "stress threshold," defined as that value of stress at which the effect on performance changes from one of facilitation to one of inhibition. Stress is formally related to time as follows:

$$S_{ij} = \frac{\bar{T}_{ij}{}^{E}}{T_j - T_{ij}{}^{U}}$$

where T_j = the total time available; $T_{ij}{}^{U}$ = the time elapsed up to but not including accomplishment of subtask i; and $\bar{T}_{ij}{}^{E}$ = the average time required for completion of all remaining essential subtasks, assuming no failures.

The probability of successful completion of a subtask increases linearly with S_{ij} from a subtask input value of $\bar{p}_{ij}$, which represents the probability that the average operator, j, can perform subtask i successfully while not under stress, until a value of unity is reached at the stress threshold. The probability then assumes the average value $\bar{p}_{ij}$, after which it decreases linearly until it reaches a value equal to the stress

threshold plus unity. At this point, it levels off at a value that is decreased from p_{ij} by an amount equal to $1-p_{ij}$. The following equations define the exact probability of successful accomplishment as a function of $\bar{p}_{ij}$, S_{ij} and the stress threshold, M_j:

$$p_{ij} = \begin{cases} \bar{p}_{ij} + \dfrac{(1-\bar{p}_{ij})(S_{ij}-1)}{M_j-1} & \text{if } S_{ij} < M_j \\ \bar{p}_{ij}(S_{ij}+1-M_j) + (M_j-S_{ij}) & \text{if } M_j \leq S_{ij} \leq M_j+1 \\ 2\,\bar{p}_{ij}-1 & \text{if } S_{ij} > M_j+1 \end{cases}$$

These exact probabilities are mapped onto binary success/failure outcomes by comparing them with a pseudo-randomly generated number R_3 uniformly distributed in the unit interval. Success is assumed if R_3 is less than p_{ij}; otherwise, failure is assumed, implying that over many trials, there will be a failure with probability p_{ij}.

Input Parameters

Values must be input for the following: T_{ij}^E, T_j, T_{ij}^U, M_j and $\bar{p}_{ij}$. As employed in the Siegel and Wolf simulation, T_j, T_{ij}, and M_j arise out of running computations on the network of task times and T_{ij}^E are derived from prior analysis of task performance time.

Model Output

As indicated above, the outputs of the model are (1) the exact probability that a given task, i, imbedded in a sequence of tasks, [i], will be performed successfully; and (2) the derived probability (equal to unity or zero) that the task will be accomplished.

Model Validation

The model has been validated by Siegel and Wolf in the course of their simulations of a wide variety of unitary and dual operator tasks and seems to represent the observed relationships between stress and performance quite well. Major experimental emphasis has centered on the search for appropriate values of M_j for any given simulation. The authors report that original values of M_j for any given simulation. The authors report that original values of this parameter were based on a derivation from the Cannon-Bard theory but that these have been modified on the basis of an empirical study by Siegel, Wolf and Sorenson, 1962.

Comment

The model successfully integrates a nominal concept of slack time (as employed, for example, in PERT) with the reasonably well validated empirical finding of an inverted U relationship between stress and performance. A particularly desirable characteristic for complex system modelling is the ability to generalize the stress-performance relation to _n_ operators performing as a team. It is likely that the model can be employed outside of the Siegel and Wolf simulation with a reasonable promise of success if an adequate basis for deriving the value of the key parameter, M_j, can be found.

Siegel & Wolf Model of Subtask Execution Time Under Stress

Reference

Siegel, A. I. and Wolf, J. J.

Description

This model, like the stress model described earlier, is employed by Siegel and Wolf to provide input into a network simulation model. Its purpose is to provide an estimate of the time required by an operator to complete a given subtask under stress when the time ($\bar{t}_{ij}$) required by the average operator and the average standard deviation of that time under non-stressful conditions are known.

The model uses the value of S_{ij} computed as described in our earlier summary and the value M_j assumed for the stress threshold. Performance is then given as

$$t_{ij} = \begin{cases} \dfrac{V_{ij}F_{ij}}{S_{ij}} & \text{if } S_{ij} < M_j \\[2ex] [(2\ S_{ij} + 1 - 2\ M_j)V_{ij} - (S_{ij} - M_j)\bar{t}_{ij}]F_j & \text{if } M_j \leq S_{ij} \leq M_j + 1 \\[2ex] [3V_{ij} - \bar{t}_{ij}]F_j & \text{if } S_{ij} > M_j + 1 \end{cases}$$

where $V_{ij} = \bar{t}_{ij} + K_{ij}\sigma ij$. ($K_{ij}$ is a random deviate, σ_{ij} is the average standard deviation of performance times, and F_j is the time required by the average operator to perform the task when not under stress.

Input Parameters

Required are the values of S_{ij}, $\bar{t}_{ij}$, M_j, K_{ij}, and F_j, the former three of which are computed and/or employed in the stress model previously identified. K_{ij} is a variable generated on the basis of random sampling from a normal distribution.

Model Output

The primary output of the model is an estimate of the time required to perform the task under various levels of stress.

Model Validation

None beyond that involved in normal execution of the Siegel and Wolf simulation model.

Comments

Comments here are similar to those made in connection with the S & W stress model. The model is modular and should be useable in any task sequence where prior empirical data regarding execution times and stress thresholds are available. A shortcoming for simulations that concentrate on "process" rather than taking a black box approach is that the model is essentially apsychological. Beyond the fact that it provides for the representation of inter- and intra-individual differences, it is not concerned with more than empirically observed relations between stress and performance.

Preyss' Bayesian Model of Learning Behavior

References

Preyss, A. E. 'A Theory and Model of Human Learning Behavior in a Manual Control Task. Sc.D. Dissertation, M.I.T., February, 1967.

Preyss, A. E. and Meiry, J. L. "Stochastic modeling of human learning behavior," Third Annual NASA-University Conference on Manual Control, NASA SP-144, 1967.

Meiry, J. L. "Stochastic modeling of human learning behavior," Fourth Annual NASA-University Conference on Manual Control, NASA SP-192, 1968.

Description

This model attempts to account for the behavior of naive subjects in a time-optimal state regulation task by means of Bayesian process of hypothesis testing. The context in which the model was first developed was bang-bang control of a pure inertia plant ($\ddot{x} = \pm U$). The model was later extended to cover other second-order plants, both stable ($\ddot{x} + kx = \pm U$) and unstable ($\ddot{x} - kx = \pm U$).

The model assumes that the human controller forms an internal model of the system state space, divided into a coarse grid, and that for each point on the grid, he formulates hypotheses about the results of a control input switch at each point. Observing the system's behavior, he generates control decisions, executes them, observes their results, and updates his hypotheses using Bayesian formulas.

Invoking Newell, Shaw, and Simon's dictum that "an explanation of an observed behavior of an organism is provided by a program of primitive information processes that generates this behavior," Preyss implemented a computer program structured in accordance with the model. Tuning the model parameters to produce switching responses similar to those of naive subjects, he succeeded in producing a program that not only converged to the proper switching strategy, but also did so in such a way that its switching responses were statistically indistinguishable from those of a sample of human subjects.

Input Parameters

Free model parameters include a priori assumptions about switching strategies, mesh intervals, assumed distributions for human decision and response times, and certain parameters characterizing the amount of reinforcement necessary to lead the controller to modify his control behavior.

Model Outputs

The model, with sufficient tuning, can reproduce sequences of control inputs statistically similar to those exhibited by naive human subjects as they refine their control strategies to converge on the optimum.

Model Validation

The model has proved sufficiently robust to be able to mimic human performance with three different sets of controlled dynamics. In each case, however, model parameters had to be adjusted to match the human subjects' data, a fact that severely limits the use of this model in a predictive role.

Comments

This model is noteworthy because it represents one of the few attempts undertaken to account for the learning behavior of naive controllers in a state-regulation task. As a descriptive model, it suffers from the need to match several context-specific parameters. It is unlikely to prove very useful for *a priori* predictions of human behavior.

The "Crossover" Model for "Tracking"

References

McRuer, D. T. & Krendel, E. S., Dynamic Response of Human Operators, WADC TR-56-524, October, 1957.

McRuer, D. T., Graham, D., Krendel, E. S. & Reisener, W., Human Pilot Dynamics in Compensatory Systems, AFFDL-TR-65-15, July 1965.

McRuer, D. T. & Krendel, E. S., Mathematical Models of Human Pilot Behavior, AGARD-AG-188, January 1974.

Description

The "crossover" model is used here as a rubric for the class of models that treat the human controller by methods of quasilinear describing function theory. The quasilinear model of the human operator consists of a describing function that accounts for the portion of the human controller's output that is linearly related to his input and a "remnant" term that represents the difference between the output of the describing function and that of the human controller.

The linear describing function portion of the quasilinear model takes on various forms depending on the precision with which one attempts to reproduce the human controller's characteristics. A fairly large body of data can be accounted for by a model of the form (McRuer, Graham et al., 1965).

$$Y_p(j\omega) = K_p \underbrace{\frac{j\omega T_L + 1}{j\omega T_I + 1}}_{\text{Equalization}} \underbrace{\frac{e^{-j\omega\tau}}{j\omega T_N + 1}}_{\text{Limitations}}$$

This describing function comprises factors related to some human limitations, namely, reaction delays (τ) and lags attributed to the neuromuscular system (T_N) and of factors used to model the human's adaptive equalization characteristics.

The most important, and perhaps most elegant, result of quasilinear manual control theory is embodied in the "crossover model" which relates $y_p(j\omega)$ to the transfer function of the controlled element $y_c(j\omega)$ by the equation.

$$y_p(j\omega)\ \ y_c(j\omega) \doteq \frac{\omega_c e^{-j\omega\tau_e}}{j\omega}$$

where ω_c is the crossover frequency and τ_e the effective time delay.

The situation with respect to the remnant portion of the quasilinear models is less well developed. The current view of remnant in quasilinear manual control theory is that, in the absence of display scanning, remnant is due largely to irreducible stochastic variation in the human operator (McRuer and Krendel, 1974, p. 65). Remnant is not error in modeling the deterministic portion of the controller's response, although such errors could contribute to remnant. Models for single-loop remnant consist of empirically obtained first-order noise spectra injected at the operator's input (McRuer and Krendel, 1974, p. 34). Fairly elaborate models for multi-display scanning have been developed and have been used to predict remnant in multi-loop situations (see, e.g., Allen, Clement and Jex, 1970).

Input Parameters

As with other control and decision theoretic models, this one assumes precise knowledge of the transfer function of the controlled element and of the disturbance inputs. The controller's describing function is determined as to form and parameters via adjustment rules that have been determined from theoretical considerations concerning closed-loop performance and an extensive empirical data base. Broadly speaking, the form of the describing function is chosen to give good low frequency response and absolute stability. The parameter values can be selected according to some optimizing criterion; however, standard practice has been to input values for quantities such as ω_c and τ_e (based on previous data).

Model Output

The model can be used to predict closed-loop system response as a function of the controlled element dynamics and input statistics. Phase and gain stability margins may be estimated, as can mean-squared errors. The model can also be used to estimate vehicle handling qualities on the basis of the describing function parameters that are determined as necessary for stabilization; in particular, the amount of lead required to stabilize a vehicle appears to be correlated with the pilot opinion of the vehicle's handling qualities.

Model Validation

The basic "crossover model" has been validated for a wide variety of single-input, single-output systems. Accordingly, it has been used in a numerous practical applications (see McRuer and Krendel, 1974). The model has also been applied and validated in more complex tasks, but to a significantly lesser extent.

Comments

Despite their unquestioned success in the analysis of a number of significant pilot-vehicle problems, quasi-linear models do suffer certain deficiencies, especially with respect to more complex tasks. One problem concerns the extension of the concepts developed for single-loop control situations to more complicated multi-input, multi-output systems. Another problem lies in selecting the parameters of the pilot describing function. These parameters are task-dependent and have been selected traditionally on the basis of verbal rules. The ability to predict remnant in complex situations within this framework is still inadequate. Because remnant can account for a significant fraction of the controller's output and increases with task complexity (McRuer, et al., 1965), a good model for predicting remnant is required to forecast performance accurately. Finally, quasilinear models are based on assumptions of stationarity and, further, are almost wholly developed for random inputs. Thus, there is no theoretical justification for extrapolating these models to time-varying, non-stationary situations or to problems involving non-random inputs. This is not to say that such extrapolations may not yield reasonable results in some instances; rather, the intent is to point out that one is on very shaky ground inasmuch as the data base underlying these models is no longer strictly applicable.

Paper Pilot

References

Anderson, R. O. "A New Approach to the Specification and Evaluation of Flying Qualities." AFFDL-TR-69-120, June 1970.

Dillow, J. D. "The 'Paper Pilot'--A Digital Computer Program to Predict Pilot Rating for the Hover Task." AFFDL-TR-70-40, March 1971.

Arnold, J. D., Johnson, R. B., and Dillow, J. D. "Pitch Paper Pilot Revisted." Proc. Ninth Annual Conf. on Manual Control, M.I.T., Cambridge, Mass., May 1973.

Teper, G. L. "An Assessment of the 'Paper Pilot'--An Analytical Approach to the Specification and Evaluation of Flying Qualities." AFFDL-TR-71-174, June 1972.

Description

"Paper Pilot" is a fixed-form, parameter optimization model of the human controller. The model was developed to predict "pilot ratings" for vehicles and tasks not believed to be covered by existing MIL-Spec flying qualities. The model incorporates a pilot-vehicle model with a form that is similar to those used in describing-function analysis. However, the distinctive feature of paper pilot is that the parameters of the model for the pilot are selected to minimize a "rating cost" functional of the form

$$R = 1 + K_o \left(\frac{\sigma-\sigma_o}{\sigma_o}\right) + \Sigma_i K_{L_i} T_{L_i} \quad (1)$$

where R is the predicted Cooper rating, σ is some measure of overall system performance (say, a linear combination of rms variations in flight path and attitude variables, T_{L_i} is the lead time constant generated by the pilot in the i^{th} control loop, and K_σ and K_{L_i} are weighting coefficients. The variable σ_o represents the desired performance level in a particular task. System performance degrades the rating only when $\sigma > \sigma_o$.

The rating expression incorporates quantities related to both performance and workload, where workload is defined in terms of the pilot lead time constants in the various loops (an interpretation that is based on crossover-model results). The amount that each performance and workload term can contribute to overall rating is constrained. This prompts the interpretation, given by Dillow (1971), that the pilot adapts his parameters to minimize a linear

combination of workload and performance, and "washes out" exceedingly poor or good performance in determining the rating. Initially "Paper Pilot" was applied to a VTOL hover task. Since then it has been used to analyze a variety of aircraft control tasks.

Model Inputs

The mathematical descriptions of vehicle dynamics, gust disturbances, etc., necessary for a control system analysis must be provided as with all quantitative manual control models. In addition, one must specify the form of pilot model (loop closures and equalization characteristics). Further, the weighting coefficients in the rating cost functional must be provided.

Model Outputs

Model outputs are predicted pilot rating (on a Cooper or Cooper-Harper scale), parameter values of the pilot model and performance measures (i.e. the standard deviations of all motion variables).

Validation

As indicated above, Paper Pilot has been used to analyze a variety of tasks. Where data have been available to validate the model, the results have been "mixed." In general, the model predicts pilot ratings with reasonable accuracy but does not do so well in predicting performance scores.

Comments

The "Paper Pilot" model with parameters selected to optimize a rating cost functional has been applied with some success. This optimizing approach replaces the verbal adjustment rules of describing function theory with a systematic procedure that should be inherently more "predictive." However, there are problems and limitations associated with the approach that have yet to be resolved.

As noted above, the model does not predict performance as well as desirable (or achievable with other models). An important reason is that existing implementations of "Paper Pilot" models do not incorporate pilot remnant. Inclusion of a remnant model may not be straightforward inasmuch as inherent pilot randomness gets confounded by "modeling error" in fixed-form models such as Paper Pilot.

Another problem is the selection of weighting coefficients in Equation (1). Although the concept of relating pilot opinion to measures of system performance and workload seems sound, general rules for choosing the K_0 and K_{Li} coefficients will be needed if the model is to be truly predictive.

Perhaps the most serious drawbacks to this technique will arise when addressing multi-output, multi-axis, multi-control problems. Then, as in the describing function approach, it will be necessary to postulate possible loop structures and the model forms for each loop. This will complicate the problem of choosing a rating cost functional, increase the possibilities of modeling error, and increase the number of parameters to be optimized. These factors will undoubtedly jeopardize the predictive capability of the techniques and will also magnify computational problems. The problems associated with computing the optimizing parameters are far from trivial even in a single loop case; for multi-loop systems - where each loop can contain several parameters - computer time could be excessive.

The Optimal Control Model

References

Baron, S. and Kleinman, D. L., "The Human as an Optimal Controller and Information Processor," IEEE Trans. Man-Mach. Syst., Vol. MMS-10, pp. 9-17, March 1969.

Kleinman, D. L., Baron, S., and Levison, W. H., "A Control Theoretic Approach to Manned-Vehicle Systems Analysis," IEEE Trans. Automat. Contr., Vol. AC-16, pp. 824-832, December 1971.

Levison, W. H., Baron, S., and Kleinman, D. L., "A Model for Human Controller Remnant," IEEE Trans. Man-Mach. Syst., Vol. MMS-10, pp. 101-108, December 1969.

Baron, S., Kleinman, D. L., et al., "Application of Optimal Control Theory to the Prediction of Human Performance in a Complex Task," AFFDL-TR-69-81, March 1970.

Kleinman, D. L. and Baron, S., "Analytic Evaluation of Display Requirements for Approach to Landing," NASA CR-1952, November 1971.

Description

The optimal control model is a stochastic, time-domain model for the human operator. It includes a model for predicting the random component of human response and is not limited to stationary control situations. It is capable of treating multivariable systems, as well as single-loop systems, within a single conceptual framework using state-space techniques. The basic assumption underlying this approach to modeling the human operator is that a highly-trained human controller will act in a near-optimal manner, subject to certain internal constraints that limit the range of his behavior and also subject to the extent to which he understands the objectives of the task.

It is convenient to consider the model for the operator as comprising the following: (i) an "equivalent" perceptual model that translates displayed variables into noisy, delayed perceived variables denoted by $\underline{y}_p(t)$. A threshold is also considered part of the perceptual process but may actually be used to model the effects of non-idealized displays. The threshold is treated by statistical linearization techniques; (ii) an information processing model that attempts to estimate the system state from the perceived data. The information processor consists of an optimal estimator and predictor and it generates the minimum-variance estimate $\hat{\underline{x}}(t)$ of $\underline{x}(t)$; (iii) a set of "optimal gains," $\underline{L}^*$, chosen to minimize a quadratic cost functional that expresses task requirements (Kleinman and Baron, 1970); and (iv) an equivalent "motor" or output model that accounts for "bandwidth" limitations (frequently associated with neuromotor dynamics) of the human and his inability to generate noise-free control inputs.

The time-delay, observation- and motor-noises and the neuromotor-lag matrix account for inherent limitations on human processing and perceptual-motor activity. Methods for choosing values for these quantities have been determined by matching experimental data and these values have been found to be generally independent of task parameters. The observation noise is a key feature of the model. It is, essentially, a lumped representation of human randomness. From the standpoint of classical quasi-linear describing function theory, the observation noise may be thought of as a model for remnant. On the basis of considerable experimentation, a relatively simple set of rules for predicting operator remnant has been found. Specifically, each display variable utilized by the operator is assumed to be perturbed by a white, gaussian, zero-mean noise process that is linearly uncorrelated with other noises and with system inputs and which has a power density level that scales with the mean squared level of the display variable.

The optimal estimator, predictor and gain matrix represent the set of "adjustments" or "adaptations" by which the human attempts to optimize his behavior. The general expressions for these model elements depend on the system and task and are determined by solving an appropriate optimization problem, and therefore, according to well-defined rules.

Model Inputs

The model inputs include representations of system dynamics, the display system, environmental disturbance commands, task requirements, and the parameters specifying operator limitations. The system dynamics comprise the linearized dynamics of

the controlled element and any dynamics associated with measurement, control, and display systems (also linearized). The equations for these dynamics are expressed in state-variable form. The quantities displayed to the human operator are assumed to be generated by linear operations on the state- and control-vectors. System and/or display may change with time, in either continuous or abrupt fashion.

Disturbance and command inputs must also be specified. Disturbances can include random gusts and quasi-deterministic inputs such as steady winds and wind-shears. Commands can include tracking inputs to the pilot that are predictable in nature but not known precisely by the pilot.

Task requirements are stated in terms of "cost weightings" associated with system variables in a quadratic cost functional. It is assumed that the operator selects his response to minimize this cost functional. The selection of cost functional weightings may be based on objective or subjective factors. For simple, single-variable control situations, good results have been obtained with a cost functional consisting of a weighted sum of system error variance plus control-rate variance. For multi-input, multi-output systems, coefficients are determined from system specifications and a knowledge of operator preferences.

Parameters related to pilot limitations are time-delay, observation- and motor-noise ratios and a "neuromotor" time-constant that reflects a limitation on operator bandwidth. Values for these parameters have been found to be relatively independent of task parameters.

Model Outputs

The solution to the defined optimization problem yields predictions of the complete closed-loop performance statistics of the system. Probability densities of all system variables (states, outputs, controls) are generated as functions of time along with mean and rms error deviations from the nominal flight path. Moreover, the densities of the pilot's estimates and estimation errors are also predicted as functions of time. All computations are performed using covariance propagation methods, thus avoiding costly Monte-Carlo simulations. However, if desired, a "sample" or simulation version of the model is possible. Given the probability density functions for the state, it is possible to compute meaningful performance statistics (e.g., one can compute the probability of a missed approach). It is important to note that the predicted probability densities are conditioned on the particular choices for system parameters. Thus, by changing particular system variables between computer runs and observing the changes in performance, the effects on performance of changes in any system quantity can be investigated systematically and relatively inexpensively.

Finally, for stationary tracking tasks, the model yields predictions of operator describing function and remnant spectrum.

Model Validation

The model has been subjected to extensive validation with very encouraging results. It has been validated in relatively simple, stationary control tasks and in more complex tasks, both stationary and non-stationary. Further validation is needed for applications that involve pre-programmed maneuvers.

Comments

The optimal control model is very general as far as manual control models go because of its time-domain, state-space formulation, its normative nature and its explicit information processing sub-model. These features have allowed models for task interference (Abstract No. 19), scanning (Abstract No. 16), decision making (Abstract No. 17) and failure detection and identification (Abstract No. 20) to be postulated and incorporated within the same framework.

Though the model has been quite successful, there remain several problem areas with respect to its use. The model is limited to linearized analysis, except to the extent that memoryless nonlinearities can be treated by statistical linearization. The choice of weightings in the cost functional is an arc involving the judgement of the analyst. It is not clear that operators, even when trained, will adopt the same or similar criteria. The model is based on an assumed state of advanced training and high motivation and its value in dealing with operators who do not fulfill these assumptions. Finally, no work has been done on applying these techniques to multi-operator situations.

Johannsen's Nonlinear Multiparameter Tracking Model

References:

Johannsen, Gunnar, "Development and Optimization of a Nonlinear Multiparameter Human Operator Model," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-2, No. 4, September 1972.

Description

This model is designed to yield highly detailed predictions of the human operator's control behavior in a wide range of time-invariant tracking tasks, including both pursuit and compensatory tracking tasks with and without prediction displays.

The model assumes that the human operator perceives the tracking error and its first derivative, his own control output and its first derivative, plus (for pursuit tracking) the forcing function and its first derivative, and (for predictor displays) the predicted error and its first derivative. Each of these signals is fed both to a control signal generating loop and to a nonlinear decision algorithm loop. In each case, the signal vector is multiplied by a weighting vector. Since the weighting vector can be different for the two loops, a total of 16 weighting parameters must be specified.

The nonlinear decision algorithm loop incorporates threshold elements that can cause the control signal generating loop to hold a particular control stick position for a period of time, regardless of changes in the other signals. Parameters of the decision algorithm loop include a threshold value, a sampling interval, and a parameter that computes the number of sampling intervals over which a constant control output should be held.

In all, there are 19 parameters that must be specified for the model. The procedure employed for determining these parameters involves both open- and closed-loop optimizations carried out using input, output, and predicted output signals recorded during simulation experiments.

Input Parameters

The transfer function of the controlled process must be input, along with values of the 19 model parameters. These parameters must either be determined from parameter optimization studies of recorded signals, or be estimated on the basis of prior studies with similar systems.

Model Outputs

The model yields predicted time histories of the human operator's control behavior for a given input signal.

Validation

Parameter values were determined for a series of experimental conditions using four different sets of controlled dynamics and displays:

(1) $H(s) = \frac{1.6}{s^2(1+0.4s)}$ pursuit tracking, predictor display

(2) $H(s) = \frac{1.6}{s^2}$ pursuit tracking, non-predictor display

(3) $H(s) = \frac{1.6}{s(1+0.4s)}$ compensatory tracking, predictor display

(4) $H(s) = 1.6$ compensatory tracking, non-predictor display

After parameter optimization, the control output signals were matched with those produced by the human subjects. The control outputs produced by the nonlinear multiparameter model matched those produced by human subjects better than the control outputs produced by a linear, describing-function model.

Comments

This model is noteworthy for its ambitious attempts to match the details of human control behavior. As a predictive tool, its use seems limited. Parameter values are reported only for a few types of systems, and little guidance is available for estimating the values that would arise with other controlled dynamics.

Phatak and Bekey's Adaptive Response Model

Reference

Phatak, A. V., and Bekey, G. A., "Model of the Adaptive Behavior of the Human Controller in Response to a Sudden Change in the Control Situation," IEEE Transactions on Man-Machine Systems, Vol. MMS-10, No. 3, September 1969.

Description

This model addresses the same problem as the Elkind-Miller adaptive response model. It assumes that following a sudden change in controlled process dynamics, the human controller proceeds through four phases of activities:

(1) a retention phase, in which he retains his pre-failure transfer function,

(2) a detection phase,

(3) an identification and modification phase, and

(4) a steady-state tracking phase using the appropriate post-failure transfer function.

The primary differences between this model and the Elkind-Miller model arise from the use of phase-plane boundaries to predict both detection and modification activities.

Specifically, the model reports that a change in dynamics has occurred whenever the system error and error rate are such that the system state exceeds the bounds of a particular region surrounding the origin. If this occurs, the model adopts a new transfer function appropriate to one of the alternative, "failed" dynamics. If the system state continues to diverge and crosses a second boundary in the phase plane, the model shifts to a second alternative, and so on.

From the description of the model published in the literature, it is not clear on what basis the order of the alternative models was chosen. The specific order used is that of decreasing probability, i.e., the most probable failure is assumed first, then the next most probable, etc. The order used also happens to be that of increasing lead-generation requirements on the human controller, in which case the "easiest" case is assumed first.

In either case, the phase-plane boundaries appear to have been constructed on an ad-hoc basis to match the controller's behavior. No rules are given for constructing these boundaries in other cases.

Input Parameters

The model requires the transfer functions of the unfailed system and of each of its alternative failed modes. The statistics of the input disturbance are also required, along with the parameters of the phase-plane boundaries that trigger the mode-switching behavior of the model.

Model Outputs

When the phase-plane boundaries are established, the model yields detailed time histories of system control inputs and system state.

Validation

The model was constructed in the context of failures in stability augmentation for the roll axis of VTOL aircraft in hover. The model was shown to reproduce the qualitative features of a human pilot's responses fairly accurately. No attempts to extend the model to other contexts were reported.

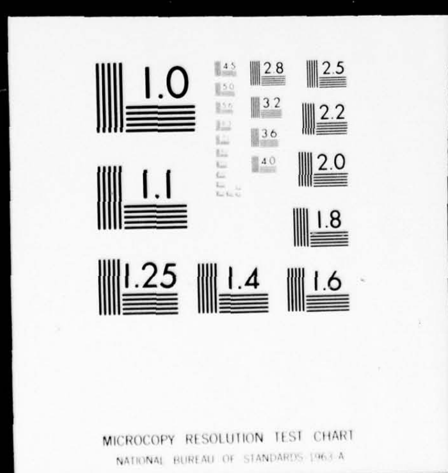
Comments

This model is noteworthy for its use of phase-plane trajectories and boundaries for the detection and identification of changes in controlled process dynamics. This approach could be useful in analyzing failure identification tasks for which data have already been recorded, but without specific rules for identifying phase-plane boundaries, the model cannot be used predictively.

AD-A038 597 BOLT BERANEK AND NEWMAN INC CAMBRIDGE MASS F/G 5/8
CRITICAL REVIEW AND ANALYSIS OF PERFORMANCE MODELS APPLICABLE T--ETC(U)
MAR 77 R W PEW, C E FEEHRER, S BARON F44620-76-C-0029
UNCLASSIFIED BBN-3446 AFOSR-TR-77-0520 NL

3 OF 4
ADA038597

ADA038597


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Veldhuyzen's Ship Maneuvering Model

References

Veldhuyzen, W.,"Ship manoeuvring (sic) under human control," Doctoral dissertation, Delft University, The Netherlands, June 1976.

Veldhuyzen, W., and Stassen, H. G. "Modelling the behavior of the helmsman steering a ship," Ninth Annual Conf. on Manual Control, M.I.T., May 1973.

Description

This model addresses a specific class of manual control problems in which the controlled process exhibits very large time constants. The specific context of the model is the helmsman's steering control of a supertanker, the dynamics of which may be described by the equation

$$T_s\dddot{\psi}(t) + a_1\dot{\psi}(t) + a_2\left[\dot{\psi}(t)\right]^3 + a_3\left[\dot{\psi}(t)\right]^{1/3} = K_s\delta(t)$$

where ψ is the ship's heading and δ is the rudder deflection angle. For large, loaded supertankers, the value of T_s can approach five minutes.

The model assumes that the helsman employs an internal estimator for deducing the system state variables, and an internal, simplified model of system dynamics. It was found that an internal model of the form

$$T_m\ddot{\psi}(t) + \dot{\psi}(t) = K_m\delta_d(t)$$

is sufficient to permit the helmsman to predict the results of his control inputs. In fact, it was shown that for these large time constants, the helmsman can barely distinguish between stable and unstable system dynamics.

Simulation studies were conducted using experienced helmsmen. Both linear (crossover) models and non-linear (phase-plane) models were matched to the experimental data. It was found that

either model could adequately reproduce the path of the ship, but that the non-linear model more closely approximated the helmsman's control inputs.

Input parameters

A mathematical description of the controlled process dynamics is required. If an accurate fit to the helmsman's control inputs is desired, sample time histories are also required in order to fit the parameters of the non-linear decision model.

Model output

The model can be used in simulation exercises to determine the ship's stability and probable course deviations during various maneuvers.

Model validation

In addition to several experiments conducted using realistic simulators, full-scale experiments were run on an actual ship under various conditions. Results showed that in each case, the helmsman's internal model exhibited a time constant very close to that of the ship itself.

Comments

This model provides a valuable case history that should be reviewed in any attempt to model human control of a system with long time constants. In very low-frequency regimes, the crossover model does not yield an adequate characterization of the controller's control behavior, and non-linear decision models must be employed.

Onstott's Multi-Axis Tracking Model

Reference

Onstott, Edward D., "Multi-Axis Pilot-Vehicle Dynamics," Proceedings of the Tenth Annual Conference on Manual Control, Air Force Institute of Technology, 9-11 April 1974.

Description

This model is designed to be used in direct digital simulation of pilot-vehicle dynamics. The model incorporates standard crossover models of pilot dynamics for each control axis, with inner- and outer-loop control of appropriate state variables. In addition, however, an "urgency function" is defined to predict the points at which the pilot switches his attention from one axis to another. This urgency function consists of a linear combination of the state variables related to each axis. It is assumed that the pilot devotes his attention to whichever axis has the highest urgency function at any particular time.

The model also includes a representation of crossfeed effects between the control inputs on the various axes. The determination of these crossfeed effects and the various crossover model gains, leads and lags is accomplished in conventional ways, but determination of the coefficients of the urgency functions requires a complicated parameter search technique.

For the primary task to which this model has been applied-a fourth-order, two-axis VTOL hovering task-stability is a major problem, and only a small set of urgency function parameters were found to yield stable model behavior. The switching behavior implied by the urgency functions was found to match very well that observed in experiments with a well-trained pilot.

Input Parameters

The controlled process dynamics must be specified, along with experimentally observed control crossfeed effects between axes.

Model Outputs

After optimization, the model yields not only predicted mean squared tracking errors and control inputs, but also predicted switching statistics between axes.

Validation

For the VTOL hover task noted above, the model successfully predicted the mean values of all state variables. It also produced histograms of time spent controlling each axis that were very similar to those observed with a well-trained pilot. The outer-loop state variable time histories produced in the digital simulation were qualitatively very similar to those observed in the experiments with the pilot.

Comments

Onstott observes that particularly in difficult, multi-axis tracking tasks, it is essential to model the interactions between the state of the system and the pilot's switching of attention from one axis to another. In the specific case studied, the system dynamics were such that only a narrow range of urgency function parameters produced stable system behavior. In such cases, mathematical optimization is likely to yield models that match very well with human performance. In other cases, where system performance is relatively insensitive to changes in the model parameters, it will not be possible to identify model parameters so precisely; under these conditions, however, the issue may be academic.

Sender's Model of Visual Sampling

References

Senders, J. W., "The Human Operator as a Monitor and Controller of Multi-Degree of Freedom Systems," IEEE Transactions on Human Factors in Electronics, HFE-5, No. 1, September 1964.

Senders, J. W., Ward, J. L., and Carbonell, J. R., "Human Visual Sampling Processes: A Simulation Validation Study," NASA CR-1258, January 1969.

Senders, J. W., Elkind, J. I., Grignetti, M. C., and Smallwood, R., "An Investigation of the Visual Sampling Behavior of Human Observers," NASA CR-434, April 1966.

Description

This model is based upon information theoretic concepts developed by Shannon and others. Fundamentally, it assumes that a human operator's fixation frequency for a particular instrument or display is dependent upon its information generation rate,

$$\dot{H} = W \log_2 \frac{A^2}{E^2} \quad \text{bits/sec} ,$$

where W is the cutoff frequency of the displayed signal, A is its amplitude, and E is the permissible rms reading error. For an observer with a fixed channel capacity, who must share his attention among several displays presenting random, uncorrelated signals with known information rates, the attentional demand of a particular instrument is calculated to be

$$T_i = 2K\, W_i \log_2 \frac{A_i}{E_i} + 2W_i\, C \quad \text{sec/sec} ,$$

where T_i is the proportion of total time that must be devoted to display i, K is a constant with dimensions of time per bit, and C (with dimensions of time per fixation) is a constant that accounts for movement time and minimum fixation time.

Using this latter relationship, the total workload placed on an ideal observer by a given set of displays can be computed. Decisions can be made regarding, for example, the effects on observer workload of adding an additional display.

For cases in which the information transmitted by the various displays is independent and uncorrelated, and in which a random scanning pattern is assumed, the probability of a scanning transition between any two displays is shown to be simply the product of their relative fixation frequencies. In general, however, it appears that scanning patterns are not random. In an aircraft cockpit, for example, a deviation on one instrument will be closely related to deviations on others, and the pilot may employ several conditional scanning patterns that depend upon the instrument readings encountered. Even in cases where the displayed information is completely independent, scanning patterns may be modified on the basis of recent readings: the observer will tend to return his attention more frequently to display showing a near-critical reading.

To address the latter issue, a model was proposed in which the controller computers the time intervals at which each signal reaches a certain probability of exceeding a threshold limit, and defers additional observations of these signals until the respective time intervals have elapsed. A possible Markov model of controller behavior was then outlined, by which improved estimates of fixation frequencies and transition probabilities could be achieved. Basically, this model assumes that the transition probabilities from display i to display j are independent of the previous sequence of transitions, although both the transition probabilities and the dwell times may be functions of both i and j.

Input Parameters

To estimate the scanning statistics of the human controller, the bandwidths and amplitudes of the various displayed variables must be specified, along with their acceptable limits. Values for the constants K and C must be deduced from data collected in a specific context.

Model Output

The model yields estimates of the fixation frequencies and durations for each displayed signal, and for the probabilities of transitions between signals.

Validation

Laboratory experiments were conducted using a set of meters displaying signals of various amplitudes and bandwidths. Subjects were required to signal with pushbuttons whenever a meter reading exceeded a certain limit. Fixation frequencies and durations

were obtained and compared with those predicted by the model. In general, the agreement was good. An attempt was made to take into account statistical correlations between displayed signals, with only limited success.

Another set of experiments was conducted in an aircraft simulator, with pilots flying various types of missions. Eye movement data were recording along with time histories of the displayed readings on each instrument. Model predictions based upon the amplitude and bandwidth of the various signals proved accurate for some pilots in some phases of flight, but not for others.

Comments

It is apparent that this model yields good predictions of scanning behavior under certain circumstances. The failure of the model to account for other results can be attributed to (1) its failure to take into account the redundancy of information obtainable from alternative sets of instruments, which permits controllers to adopt quite different alternative scanning strategies under various conditions, and (2) its failure to take into account the interactions between control behavior and instrument sampling, which permit the controller to estimate changes in the displayed signals. A closely related problem is that particular signals become more or less critical during different maneuvers, an effect that is largely ignored by this model.

Smallwood's Instrument Monitoring Model

References

Smallwood, R. D., "Internal Models and the Human Instrument Monitor," IEEE Transactions on Human Factors in Electronics, Vol. HFE-8, No. 3, September 1967.

Description

This model grew out of previous instrument monitoring studies involving Senders, Elkind, Grignetti, and Smallwood. The task under consideration involves a human controller monitoring the readings on four meters, which are driven by signals of different amplitudes and bandwidths, and signalling whenever any meter exceeds a certain threshold.

The model assumes that the monitor constructs an internal model of the processes being monitored, and that at any instant he can estimate the probability density functions of the amplitude of each signal. This p.d.f. will, of course, depend on the time elapsed since his last reading. Immediately after an instrument is read, its p.d.f. is an impulse; i.e., the monitor is assumed to have perfect knowledge of its amplitude. As time passes, however, the variance of the p.d.f. increases and its mean approaches the average amplitude of the signal. As this occurs, the probability that the signal has exceeded the threshold tends to increase. The model assumes that each scanning transition is to the instrument for which the probability of exceeding the threshold is greatest.

The model further assumes that a "dead time" of 0.1 second is required to shift attention between two instruments and that the time required to read an instrument is inversely related to its distance from the threshold. The precise form of this relationship must be deduced from data collected in the appropriate context.

Input Parameters

A statistical description of each displayed signal is required, along with its respective threshold. Overall fixation frequency must be matched to that observed in the appropriate context.

Model Output

The model yields predictions of the relative fixation frequency and duration of fixation for each instrument, as well as predictions of the average transition probabilities between the instruments.

Validation

The model predictions were compared with data obtained by Senders in another experiment (see Abstract No. 12). parameter was adjusted to match the observed overall scanning frequency, the model output matched the experimental results fairly well. A second-order version of the human controller's model (i.e., one that took signal velocity into account along with signal amplitude) produced a better overall fit than a first-order version.

Comments

This model represents an improvement over Sender's model in that the effect on scanning behavior of the proximity of a displayed signal to its critical threshold is explicitly taken into account. No attempt was made to model the effects of possible cost differences in threshold passages of different signals. The interactions between the displayed signals and the operator's control inputs were also excluded from consideration.

Carbonell's Queueing Model of Visual Sampling

References

Carbonell, J. R., "A Queueing Model of Many-Instrument Visual Sampling," IEEE Transactions of Human Factors in Electronics, Vol. HFE-7, No. 4, December 1966.

Senders, J. W., Carbonell, J. R., and Ward, J. L., "Human Visual Sampling Processes: A Simulation Validation Study," BBN Report No. 1681, June 1968.

Description

This model for a pilot's instrument scanning behavior is based upon queueing theory concepts. The observer is modeled as a single-channel server that can attend to only one instrument at a time. It is assumed that at each step in his scanning process, the observer tries to minimize the risk that an instrument to which he is not attending will exceed some threshold reading. It is further assumed that the time involved in reading each instrument is constant (approximately 1/3 second), and that whenever a longer observation time is noted, it is because the observer has chosen to look at the same instrument again to minimize total cost.

The total cost of not looking at any instrument is defined to be

$$C(t) = \sum_{i=1}^{M} \frac{C_i \, P_i(t)}{1 - P_i(t)} \quad ,$$

where M is the total number of instruments, C_i is the cost associated with exceeding the established threshold for instrument i, and $P_i(t)$ is the probability that instrument i will exceed its threshold at time t. The total cost of looking at instrument j at time t is

$$C'(t) = C(t) - C_j \, P_j(t) \quad .$$

The model assumes that the pilot chooses the instrument j that will minimize C'(t) at any instant. Each instrument is treated as a member of a set competing for the pilot's attention. The nonlinear dependence of cost on the probability of an instrument's exceeding a threshold insures that every instrument will eventually reach a point where it must be attended to.

A significant feature of the model is that the observed signals are not assumed to be gaussian with zero mean. This permits the model to deal with signals that are modified by feedback loops involving the pilot's control actions.

The model is implemented as a computer simulation, which can be exercised using different assumptions regarding cost thresholds, signal characteristics, etc.

Input Parameters

To exercise the model, one must specify the statistical characteristics of each displayed signal, the costs of exceeding given thresholds on each display, and the thresholds below which each instrument reading is ignored.

Model Outputs

The model yields a time sequence of instrument fixations, which may be abstracted to identify patterns of scanning activity.

Validation

This model was applied to eye movement data obtained from several experienced pilots in an instrument flight simulator. Using estimates obtained from the pilots of the relative cost of various instrument deviations in different phases of flight, a computer simulation of predicted scanning behavior was run. The results of this simulation agreed quite well with the experimental data - much better than did simpler scanning models based only on instrument bandwidths. The model was found to be particularly valuable as signal variances and task difficulty increased, conditions in which greater scanning aperiodicity is usually observed.

Comments

This model represents a significant advance in modelling instrument scanning behavior. It includes not only differential costs for deviations on various instruments, but also a method for incorporating the effects of the pilot's control activities. The model does not attempt, however, to take into consideration cross-coupling among instruments, though Carbonell asserted that such an extension is possible. The flexibility and power of this model is obtained at the cost of considerable analytical complexity. It appears that the complexities of scanning behavior can adequately be studied only through extensive simulation.

Frequency Domain Scanning Model Using Quasilinear Describing Function Theory

References

Allen, R. W., Clement, W. F., and Jex, H. R. Research on display scanning, sampling, and reconstruction using separate main and secondary tasks. NASA CR-1569, July 1970.

Hoffman, L. G., Clement, N. F., and Blodgett, R. E. Further examination of pilot instrument scanning data and development of a new link value estimator. Proceedings of the Ninth Annual Conference on Manual Control, M.I.T., Cambridge, Mass., May 23-25, 1973.

McRuer, et al. A systems analysis theory for displays in manual control. Systems Technology Inc. Report No. 163-1, June 1968.

Description

This model treats the processes involved in scanning, selecting, sampling and reconstruction of signals displayed on a multi-instrument panel by adding a "perceptual" block to the classical quasi-linear model of the pilot. This added block is stipulated to account for the effects of scanning and sampling, such efforts be defined as the ratio between behavior under continuous, full-foveal tracking and the actual sampled tracking, in each of the multiple loops of control.

The model is based on the following fundamental assumptions: 1) scanning and sampling in closed loop, multi-variable control can be accounted for by models that are extensions of quasi-linear models used for non-scanning cases; 2) scanning behavior is stable in a statistical sense; 3) sampling of a given display is assumed to be almost periodic with appreciable statistical fluctuations that randomize the data; and 4) high frequency effects of scanning sampling and reconstruction give rise to a broadband "sampling" remnant that can be modelled as an injected noise at the pilot's input (i.e., "observation noise").

The resulting model is comprised of a quasi-linear random input, "perceptual describing function" which multiplies the basic describing function and a broadband sampling remnant which adds to the basic remnant. The perceptual describing function consists of an attenuation factor and an equivalent sampling and reconstruction delay. The spectrum of the injected sampling is proportional to the mean-squared value of the signal being sampled and, for frequencies well below the average sampling frequency, is approximated by:

$$\Phi_{nn_s}(\omega)\Big|_{\omega<<\omega_s} \approx \frac{\bar{T}_s\,\bar{x}^2\left(1-\frac{T_D}{\bar{T}_s}\right)\left(1-\frac{T_o}{\bar{T}_s}\right)}{\pi\left[1+\left(\frac{\omega T_D}{2}\right)^2\right]}$$

where x is the observed signal, $\bar{T}_s$ the average sampling interval, T_D the dwell time and T_O the minimum sampling interval.

Input Parameters

Application of the model requires all information necessary for a manual control analyis, i.e., a complete system description and a model for manual control. The system description must include mission description and criteria, vehicle dynamics, disturbances and commands in a suitable form. The manual control model is the multi-loop quasi-linear describing function model modified to include the "perceptual" describing function.

Model Outputs

The outputs of the model are those of a standard manual control analysis plus quantities related to display design and utilization such as fixation frequencies and durations, scanning patterns, scanning workload metrics and preferred presentation and arrangement.

Model Validation

The model has been used to analyze a B707-320 in ILS approach and predicted display arrangements agree quite well with those adopted for an American Airlines aircraft (McRuer, et al). However, detailed validation for such a complex task is not available. The model has also been tested in carefully structured two display cases (Allen, Clement and Jex) with encouraging results.

Comments

The strengths of the model are those of the elemental models. However, the drawbacks of the individual models may be compounded by the syntheses.

Optimal Control Scanning Model

References

Baron, S. and Kleinman, D. L., "The Human as an Optimal Controller and Information Processor," IEEE Trans. Man-Machine Syst., Vol. MMS-10, pp. 9-17, March 1969.

Baron, S., et al., "Application of Optimal Control Theory to the Prediction of Human Performance in a Complex Task," AFFDL-TR-69-81, March 1970.

Description

The optimal control scanning model is a method for including scanning constraints in the optimal control framework. The model is closely related to the concepts in Levison's attention-sharing model. The main difference is that the scanning model accounts for perceptual losses associated with large eye movements.

The basic idea is to assume that the observation noise process depends parametrically on the fixation point of the eye, represented mathematically by the parameter ω. A scanning strategy is defined as a method for picking this fixation point at different time instances, i.e., a procedure for choosing $\omega(t)$. It is assumed that the pilot's fixation point cannot be changed instantaneously so that there is a finite transition time t_o, associated with each such change. The transition time also includes time lost due to the apparent loss of perception that occurs immediately before and after each eye movement. Note that $\omega(t)$, so defined, is piecewise constant and, over any time interval, may be described by a scan sequence of the form (say) $\{1,0,2,0,1,.....\}$.

When the fixation point of the eye varies with time, the observation noise covariance matrix $\underline{V}_\omega$ will also vary with time. However, for each value of ω there is a corresponding constant observation noise covariance matrix. If $\underline{V}_i$ is the covariance matrix corresponding to ω=i, then the diagonal element in the j^{th} row, $V_i(j)$, is the noise-covariance associated with viewing the j^{th} displayed variable while fixating foveally display i. Note that each of the matrices $\underline{V}_i$ have elements corresponding to viewing each of the displayed variables. Some of these elements will correspond to foveal viewing and others to peripheral viewing. Put another way, all the displayed variables are assumed to be observed continuously, but, in general, the signals on the "fixated" display will have lower noise levels associated with their observation; this influences the scanning strategy. In formulating the method for predicting

scanning behavior, it is assumed that the matrices V_i, $i \neq 0$, are given. The elements of the matrix $\underline{V}_O$ are assumed to be infinite, i.e., that nothing is seen while the operator is switching points of fixation.

It is also assumed, without loss of generality, that the scanning strategy is periodic with a fixed, but arbitrary, scan period T. It is then possible to show that for a given scanning strategy, ω, the quadratic performance criterion of the optimal control model depends implicitly on ω. Therefore, consistent with the optimality hypothesis, it is assumed that the pilot selects his control ($\underline{u}$) and scanning strategy (ω) to jointly minimize the performance criterion.

Model Inputs

In addition to those inputs necessary for the optimal control model, one must provide estimates of the average transition times between instruments and estimates for observation noise covariances for peripheral viewing of displays. (The simplest assumption is that nothing is seen in the periphery.)

Model Outputs

In addition to the standard outputs available from the optimal control model, one obtains estimates of average scanning behavior. The quantity t_i can be identified with mean observation times for display i (for all i). The "predicted" average scan period is the optimal value of T. It will be noted that errors in the prediction of scan period will inevitably result in errors in "predicting" mean observation times. Therefore, it is better to consider the fraction of time spent fixating display i foveally ($f_i = t_i/T$) instead of the mean observation time on that display (t_i). The fractions of fixation and the average scan period are quantities that can be compared directly with values obtained from measured eye-movement data. If it is assumed that instrument transitions form a Markov chain in which the observer draws at random from the set of displayed signals with probabilities equal to the fixation fractions f_i each time a display transition is to be made, then transition link probabilities and average dwell times can be computed. (The fixation fractions are then identified with fixation probabilities, which is tantamount to assuming ergodicity of the scanning process.)

Validation

Validation results for the optimal scanning models are very meager, being limited to a demonstration that operator performance in a VTOL hover task with two separated displays could be reproduced.

Comments

The advantages of this approach to scanning prediction are several. The sampling model is such that the human's monitoring behavior depends upon the control requirements and control actions in an explicit way. Control and scanning strategies are determined at essentially the same time; one avoids the inherently iterative and judgmental procedures involved in a process that requires the specific loop topology to be known before computing scanning parameters. Also, one avoids separate assumptions concerning reconstruction of the sampled signal; the Kalman estimator performs this function, as in the no-scanning case. Other advantages include the straightforward manner in which peripheral vision may be considered and the lack of any restrictive assumptions concerning coupling of signals on various display indicators.

The principal difficulty associated with the approach would appear to be the computational burden in solving the optimal sampling problem; it should be noted, however, that solving the optimization problem may not be more expensive than the computational procedures associated with other methods for predicting scanning in closed loop control tasks. In the two-display case numerical search techniques are adequate and not too expensive. For more than two displays, search methods are not too practical and a suitable optimization algorithm is required.

Control Theory Model for Decision Making

References

Levison, W. H. and Tanner, R. B., "A Control-Theory Model for Human Decision Making," NASA CR-1953, December 1971.

Levison, W. H., "A Control Theory-Model for Human Decision Making," Proc. IEEE Conference on Systems, Man and Cybernetics, Anaheim, California, October 1971.

Description

This is an extension of the optimal control model of the human operator to tasks involving continuous detection and/or decision-making. The model is identical to the optimal control model in terms of its assumptions concerning information processing. Thus, it is assumed that, because of inherent limitations, the operator perceives a noisy, delayed version of the displayed variables. The perceived data are then "processed", via an optimal estimator/predictor combination, to generate a minimum variance estimate, $\hat{\underline{x}}(t)$, of the system state vector and the covariance of the error in that estimate, $\underline{\Sigma}(t)$. The pair $(\hat{\underline{x}}(t), \underline{\Sigma}(t)$ is a sufficient statistic for testing hypotheses about the state of the system.

The model assumes that the operator is an optimal decision-maker in the sense of maximizing expected utility. This strategy is then applied to the problem of deciding whether or not a signal, corrupted by noise, is within certain prescribed tolerances. For equal penalties on missed detections and false alarms, this rule reduces to one of minimizing the expected decision error. The resulting decision rule is, simply, a likelihood ratio test.

Model Inputs

Model inputs include those necessary to apply the optimal control model with the exception of motor-noise and the quadratic cost functional. Instead of the latter, one must specify the decision cost, i.e., the utilities for various correct and incorrect decisions.

Model Outputs

Many measures of decision or detection performance can be predicted. Thus, one can compute the decision error probabilities or other measures such as the distribution function for failure detection time. It should be emphasized that the model can be used to explore the effects on decision performance of changes in system or display parameters.

Validation

Experimental results have been compared with model predictions for the following task situations (1) single decision task; (2) two decision tasks and (3) concurrent manual control and decision tasks. Using fixed values for model parameters, we can predict single-task and two-task decision performance scores to within an accuracy of 10 percent. Interference between concurrent decision and control tasks is revealed, but the results of this experiment did not allow a conclusive test of the predictive capability of the model.

Comments

One can consider decision making as being comprised of two processes, the generation of a decision function (i.e., the process of converting actual observations into one or several numbers) and the formulation of a rule or law which uses the output of the decision function to decide between hypotheses. In this model, the information processing structure of the optimal control model generates the decision function and the decision rule is essentially that of a Baysian decision maker. It is clear that other decision rules may be formulated (prescriptive or descriptive). These will lead to somewhat different models but all will incorporate an explicit information processing model that is based on the optimal control model.

The decision making model has only received limited validation. Moreover, there were several methodological problems in the experiments used for validation. These are discussed in the cited reference. Thus, the model needs further verification before it can be used with confidence.

The Elkind-Miller Model for Human Controller Adaptation to Sudden Changes in Controlled Process Dynamics

References

Elkind, J. E. and D. C. Miller, "Adaptive characteristics of the human controller of time-varying systems," Bolt Beranek and Newman Inc., Cambridge, Massachusetts, Report No. 1360 (April 1966); also AFFDL-TR-66-60, Wright-Patterson Air Force Base, Ohio, December 1967.

Miller, D. C., "A Model for the adaptive response of the human controller to sudden changes in controlled process dynamics," S. M. Thesis, M.I.T., June 1965.

Elkind, J. I. and D. C. Miller, "The adaptive response of the human controller to sudden changes in controlled process dynamics," IEEE Transactions on Human Factors in Electronics, HFE-8, September 1967, pp. 218-223.

Description of Model

This model addresses the problem of human adaptation to sudden changes in controlled process dynamics in a compensatory tracking task. It attempts to predict when a human controller will first become aware of an unanticipated change in dynamics (such as the failure of a stability augmentation system), and how he will change his control behavior to compensate for the transition.

The model assumes that the controller is well-trained and familiar with all of the possible changes in the dynamics that can occur. It postulates that he continuously monitors the change in the system error rate, $\Delta\dot{e}$, that results from his control inputs, and compares this change with the predicted change that he expects on the basis of his experience with the system dynamics. If the difference between the observed and expected changes in error rates exceeds a criterion value, the model reports that a change in system dynamics has occurred.

The model then enters a Bayesian analysis phase in which the controller compares the observed change in error rate with that expected under alternative hypotheses concerning the system change that has occurred and selects that set of alternative

dynamics that best accounts for the observed results. Usually, only one candidate will match adequately, but if not, identification is postponed and additional observations are accumulated until an unequivocal choice can be made.

The model asserts that the controller then carries out a rapid, often pre-programmed series of control movements to null accumulated errors, and finally conducts a "fine-tuning" process to optimize his transfer function for the new controlled process dynamics.

Input Parameters

The model assumes precise knowledge of the transfer function of the "normal" controlled process and each of the alternative "failed" processes. It also requires statistical descriptions of the disturbance signal being tracked and of the resulting error signal. The steady-state tracking characteristics of the human controller are assumed to be consistent with the "crossover model" (see Abstract No. 5).

Model Outputs

For well-defined tracking tasks, the model predicts the time required to detect and respond to various changes in controlled dynamics. With sufficient knowledge of specific tracking situations, the probability of an erroneous response can also be predicted.

Model Validation

The model has been applied to data obtained in compensatory tracking tasks conducted in the laboratory using K, K/s, and K/s^2 controlled dynamics. Transitions studied included gain increases, gain decreases, and gain reversals, as well as changes in the order of the controlled dynamics (e.g., K/s^2 to K/s). Experiments were conducted with up to 18 possible transitions, and also with transient disturbances designed to mimic the error waveforms of an actual transition.

The model successfully predicted the response times of the various transitions and, in many cases, accounted for errors in identifying the transition that had occurred.

Comments

The utility of this model lies primarily in its capability to account for the detailed behavior of well-trained human subjects responding to sudden changes in controlled dynamics. It represents a "best case" model of this behavior; in most realistic tracking situations, however, the controller is not aware of all possible malfunctions and is not well-practiced in responding to them. Therefore, this model is of limited usefulness in predicting the consequences of system failures.

Levison's Model for Attention Sharing and Workload

References

Levison, W. H., Baron, S., and Kleinman, D. L., "A Model for Human Controller Remnant," IEEE Trans. Man-Machine Syst., Vol. MMS-10, pp. 101-108, December 1969.

Levison, W. H., Elkind, J. I., and Ward, J. L., "Studies of Multi-Variable Manual Control Systems: A Model for Task Interference," NASA CR-1746, May 1971.

Levison, W. H. and Tanner, R. B., "A Control-Theory Model for Human Decision Making," NASA CR-1953, December 1971.

Description

Levison's model for attention-sharing is intended to apply to situations (such as continuous manual control) in which the human is required to operate as a continuous processor of information. The model is further restricted to situations in which task interference can be assumed to be of central origin. (Visual scanning effects can be treated in an analogous fashion.) This model serves as a theoretical framework for including disruptive effects of workload on operator performance.

Levison's model follows directly from his model for controller remnant or randomness. That model states that each variable "perceived" by the operator is perturbed by a white, gaussian noise process with autocovariance given by

$$V_i(t) = \pi P_i \sigma_i^2(t) \qquad (1)$$

where $\sigma_i(t)$ is the variance of the i^{th} perceived variable and P_i is the "noise/signal ratio" and has units of normalized power per rad/sec. Numerical values for P_i of 0.01 (i.e., -20 dB) have been found to be typical of a variety single-variable control situations. The relative invariance of P_i with control task parameters suggests that the basic observation noise defined by Equation (1) represents a processing limitation of the human operator.

The model for attention sharing (or task interference) assumes that if two or more display variables are to be processed, the noise/signal ratios may deviate from their nominal values. (No deviation is assumed when position and rate information are obtained from the same display or when the display is integrated.) Specifically, the model for interference assumes that as the human devotes less "attention" to a given displayed variable the noise on that variable increases proportionally. Thus, if P_o is the noise-ratio

when "full attention" is devoted to a task, then when the subject is forced to pay less than full attention to the i^{th} variable, the effective noise/signal ratio is

$$P_i = \frac{P_o}{f_i} \tag{2}$$

where f_i is the fraction of attention, and $0 \leq f_i \leq 1$, $\Sigma_i f_i = 1$. Clearly, because of the increase in noise, attention sharing will degrade performance reliability.

Equation (2) can be used to help predict the effects of attention-sharing on performance. To predict what happens on a specific task when only partial attention can be paid to it, (2) in conjunction with (1) is used to determine the appropriate observation noise covariance. Then the optimal control model is used to predict pilot behavior and overall system performance. The optimum allocation of attention is determined and provides the basis for prediction of the other measures of interest.

The model of task interference lends itself straightforwardly to the prediction of the amount of "workload" associated with a given task. The "workload index" (WI) is defined as the fraction of the controller's capacity that is required to perform a given task to some specified, or criterion, level of performance. This metric can be predicted quantitatively with the existing implementation of the optimal-control model. The procedure is identical to that for predicting task interference: once the model is "calibrated" for single-axis behavior (either by doing a simple experiment or by using nominal values of parameters that match previous data), a curve of performance score versus observation noise ratio is obtained. By relating the observation noise ratio to fraction of capacity, a quantitative value of workload may be determined.

Model Inputs

Inputs to the model are the base noise/signal, P_o, and the definition of those variables (or tasks) that are assumed to be mutually interfering. Of course, all inputs necessary for applying the optimal control model must also be specified.

Model Outputs

In addition to those system performance measures provided by the optimal control model, the model for attention-sharing yields the optimal allocation of attention to the various tasks. If performance specifications are available, a prediction of workload is also made.

Validation

The model for task interference has been validated in basic two-axis and four-axis tracking tasks. It has also been validated, but to a lesser extent, in a two-variable decision-making task. The effects of interference on total performance appear to be predicted with excellent accuracy. However, sub-task scores do not agree so well which suggests that the subjects' allocation of attention may deviate from that predicted by the model.

Comments

There are several basic problems associated with application of the task interference/workload models. First, there is the necessity for specifying a "reference" noise/signal ratio P_o. Ideally, P_o should correspond to full attention. However, conducting an experiment to measure P_o does not seem possible. Thus, one uses the value of P_o obtained in standardized laboratory tracking tasks in which the subjects are well-trained and highly motivated to minimize tracking error. It is assumed that this value of P_o corresponds to a "demand" that would correspond to a high workload outside the laboratory. If one is interested in the relative change in workload from one situation to another, as is often the case, then the value of P_o is not too

A second problem involves solving for the optimum allocation of attention. This is a non-trivial computation problem. Inasmuch as it is often the case that reasonably large deviations from optimal allocation fractions do not result in seriously degraded performance, it is frequently desirable to choose f_i's to reflect some hypothesized allocation of attention. This leads to a descriptive model of attention-sharing rather than a prescriptive one. The fact that overall performance does not suffer much when the attention-split deviates from optimal is one reason why sub-task performance is not so well predicted by the model.

A third difficulty arises in deciding whether variables or tasks do, in fact, interfere. This is especially so when display information is presented in an integrated format, e.g., in a pictorial display. At present, sufficient data do not exist to allow one to determine theoretically the degree of integration of any given display. However, by obtaining model predictions for the alternative assumptions of complete interference and no interference, the range of performance to be expected can be predicted.

Finally, it should be mentioned that the model for attention-sharing did not appear to hold when attempting to predict interference between a tracking task and a decision-making task.

Gai's Failure Detection Model

References

Gai, E. G. and Curry, R. E., "Failure Detection by Pilots During Automatic Landing; Models and Experiments," Proc. of Eleventh Annual Conference on Manual Control, NASA TMX-62,464, Ames Research Center, Moffett Field, California, May 1975.

Gai, E. G., "Psychophysical Models for Signal Detection with Time Varying Uncertainty," Ph.D. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, January 1975.

Description

This is proposed as a model of the pilot as a monitor of instrument failures during automatic landing. The model incorporates observation noise, a Kalman estimator and a decision rule based on sequential analysis. The observation noise is selected according to Levison's task interference model.

The Kalman estimator is used to generate the decision function. In particular, the estimator generates estimates for the state, $\hat{x}(t)$, the measurements, $\overline{y}(t)$, and a measurement residual

$$\varepsilon(t) = y(t) - \hat{y}(t)$$

where $y(t)$ is the "noisy" measurement. In addition, covariances for these quantities are also available from the filter. Under normal operating conditions the residual is a zero mean white gaussian noise process. It is assumed that in a failure, a deterministic is added to the measurement. In that case the residual is still white and gaussian with the same variance as in the unfailed case, but it has a non-zero mean. Thus, Gai chooses as his decision function

$$\lambda(m) = \sum_{i=1}^{m} \{\varepsilon(t_i) - \theta_{1/2}\}$$

where $\theta_1 > 0$ is the mean value associated with a failure.

The decision rule that is used is a modified form of the classical sequential analysis rule, namely

$$\text{if } \lambda(m) \geq \frac{\ln A_1}{\theta_1} \quad \text{choose "failure"}$$

$$\text{if } 0 \leq \lambda(m) \leq \frac{\ln A}{\theta_1} \quad \text{take another observation}$$

$$\text{if } \lambda(m) \leq 0 \quad \text{"reset" } \lambda(m) \text{ to zero}$$

The value of A_1 is related to the desired false alarm and miss probabilities by

$$A_1 - \ln A_1 - 1 = -[\ln A + (A-1) \ln B/(1-B)]$$

where

$$A = (1 - P(MS))/P(FA) \quad ; \quad B = P(MS)/(1-P(FA))$$

Gai claims that classical sequential analysis cannot be applied to the failure detection problem because the basic assumption of that theory is that the same mode exists during the entire period of testing.

Model Inputs

System and disturbances must be given as in the optimal control model. In addition, observation noise/signal ratio for the human operator, the parameter, θ_1, designating the mean of the failed process, and the "desired" decision performance parameters, P(MS) and P(FA), must be specified.

Model Outputs

The model predicts the mean and standard deviation of the detection time for failures. These can be determined as a function of system parameters.

Validation

An experiment to test the validity of the model was conducted. In this experiment subjects had to detect failures in the glide-slope and airspeed indicators during a simulated landing in a Boeing 707 fixed base simulator. The results showed that good agreement between predicted detection times and experimental data.

Comments

The information processing part of the optimal control model is incorporated in this model except that the time-delay (and, therefore, the optimal predictor) are ignored. The model differs from Levison's model for decision-making in the choice of filter output for a decision function and in the decision rule.

It should be noted that the model treats only instrument failures that are evidenced by changes in the mean value of the display variable.

Human Operator Simulator

References

Strieb, M. I. The Human Operator Simulator: Volume I Introduction and Overview Analytics Technical Report 1117-1 August 1975.

Description

The goal of HOS is to provide a computer simulation of human performance in a goal-oriented, task-processing environment. It is used in conjunction with other programs such as the Human Operator Procedures (HOPROC), Human Operator Data Analyzer/Collator (HODAC), and the HOPROC Assembler/Loader to provide the capability for full simulation.

Four High Level Functions are defined:

1. The decoder, which provides the translation from task requirements to human performance functions.

2. The multiplexor, which sets priorities for the order in which task requirements will be sequenced, based on an index of criticality.

3. The estimator, which is the sensing-remembering core of HOS and assigns time costs to the elemental human performance operations undertaken. It includes a memory module that is based on a strength model of memory performance having a short-term and long-term memory capability. The memory module is the major probabilistic element in HOS. The estimator also includes a model of the operator's physical and biomechanical operations such as eye movements, limb movements, etc.

4. The banker, which accumulates time costs and serves to interface HOS with the simulated hardware components of the system.

In order to build a man-machine simulation in HOS, the designer starts from a task analysis and writes in HOPROC a procedural description of the activities to be performed. Where the translation from procedures to human performance functions has been previously defined for generic activities like instrument reading, only parameter values must be specified.

However, if new procedures are encountered it is necessary to provide the translation from procedure to human performance functions and, where necessary, new human performance function definitions.

In terms of the vocabulary used in this report HOS represents a bottom-up modelling approach that carries the detail down to the level of specifying the elemental human performance functions out of which more molar task-level behavior is built.

Input Parameters

In the extreme case the model user must specify the operational procedures, the definition of the relationships between tasks and human performance functions, the description of the human performance functions and the parameter settings that instantiate the models at each of these levels for the actual task being represented. In practice the user may take advantage of the accumulated catalog of reprsentations at these different levels and specify only the specific parameters relating to his simulation. However, where new representations are required, descriptions of the models themselves must also be provided.

Model Output

The HODAC program makes possible user specification of any desired output variables and statistical summaries of them within the limits of the level of detail specified in the models themselves.

Model Validation

HOS takes a cumulative incremental approach to model development and validation. Validation of HOS may be undertaken in successively larger units. We are aware of validation to predict meter reading and scanning performance, for example. For this purpose, the models required for this subset of tasks have been implemented and tested against real data. When these models have been tuned, the underlying human performance components are available to predict larger units of behavior of which instrument reading is a subtask, and so forth. We believe that this is an admirably systematic way to proceed, but, thus far, the conduct of validation studies of this kind has been limited.

Comments

The cumulative incremental approach to development and validation that the underlying structure of HOS makes possible is an important feature of the methodology. A second important contribution is the insightful and careful way in which the various components of human performance are being modularized and interrelated. As a result, many of the aspects of human performance fall out as a consequence of these logical relationships, thereby reducing the complexity of task modelling requirements.

The ultimate success of HOS will depend on two factors: (1) the generality with which individual human information processing operations can be modelled; and (2) the ability to provide an integrative structure for combining elements of human performance at one level and tasks at another level so that the model is predictive of system-level performance.

Information Measurement

References

Fitts, P. M., Posner, M. I., Human Performance, Monterey, Calif., Brooks/Cole 1967.

Garner, W. R., Uncertainty and Structure as Psychological Concepts, New York, Wiley, 1962.

Gross, R. L., Lindquist, O. H., Peterson, J. R., Blanchard, J. D. "The Application, Validation and Automation of a Method for Delineating and Quantifying Aerospace Flight Crew Performance." Symposium and Workshop on Quantification of Human Performance University of New Mexico, August 1964.

Pew, R. W., "Recent Psychological Research Relevant to the Human Factors Engineering of Man-Machine Systems." Proceedings National Electronics Conference. Chicago, Ill., Vol. 21, 1965.

Senders, J. W., "The Human Operator as a Monitor and Controller of Multi-degree of Freedom Systems," IEEE Transactions on Human Factors in Electronics, HFE-5, 1964, p. 2-5.

Teichner, W. H., Krebs, M. J., "Laws of Visual Choice Reaction Time," Psychological Review, 81, 1974, 75-98.

Description

Although not strictly a model, the concepts of information measurement may be used to formulate models for a variety of human information processing activities. The fundamental finding is that when there is a one-to-one mapping of stimuli to responses, choice reaction time is a linear function of the average information (uncertainty reduction) when responding to a stimulus drawn from a well-defined stimulus set whose probabilities of occurrence are known. In the discrete case, the average information is defined as

$$H_{ave} = -\sum_{i-1}^{n} p_i \log_2 p_i. \tag{1}$$

and

$$RT = a + b\,H_{Ave}. \tag{2}$$

This finding holds for varying information by either varying the number of stimuli in the set, or by varying the relative frequencies of occurrence of particular stimuli in the set and then varying the first-order sequential probabilities that a particular stimulus will follow any other stimulus.

This representation works best when the stimuli are separate lights, numbers, or other unitary events that either occur or do not occur. However, the method of modelling may also be applied to predict the time to read a visual display. In this case, one must quantify the probability distribution of alternative display readings using either the discrete or continuous version of the information metric. The discrete version is given above. The continuous version is defined by

$$H_{AVE} = \omega \log_2 \frac{A^2}{E^2} \quad \text{where}$$

ω = bandwidth of time variation in meter reading.

A = rms meter reading amplitude

E = rms amplitude of acceptable error

If an operator is required to monitor several such displays, the workload associated with reading a set of such instruments may be estimated by deriving the information load of each one and relating the sum of individual loads to an *a priori* estimate of the operator's channel capacity (see Senders, 1964).

Input Parameters

In order to compute H, the probability distribution over the stimulus set is required. It may be known precisely or estimated. In order to predict response time, the intercept and slope, a and b of Eq. (2), are required. The value of parameter a will depend on the nature of the response required. If no overt movement is required, a = 200 msec. will provide an adequate approximation. The parameter b depends on the extent of practice and on the stimulus-response compatibility of the action to be undertaken. In the absence of more detailed information, a slope of .33 sec/bit or a 3 bit per sec information rate may be assumed as a rule of thumb. Teichner and Krebs, (1974) summarizes parameter values for a wide variety of laboratory situations.

Model Outputs

The model as presented here predicts response time. However, a unique feature of the information measure is its implicit assumption of a relation between speed and accuracy of performance. For a given set of conditions and at a given level of practice, rate of information transmission has been shown to remain relatively constant; thus, if the operator increases his speed of performance, the transmitted information decreases proportionately. The computation of information transmitted requires detailed information about the distribution of responses and errors. If those data were available a priori, predictions of information rate, a measure combining speed and accuracy would be possible, but usually these data are only available after test.

Model Validation

The basic relation between information and response time has been extensively validated in the laboratory, (see Teichner and Krebs, 1974). In one application (Gross, et al, 1964) these methods together with other models for prediction of movement times derived from MTM principles were used to predict workload for one phase of the Mercury spacecraft mission. Predictions were within 10-15% of values obtained from live mission simulations.

Comments

Although the popularity of the information metric among the research community fourished in the 1950's and early 60's and has decayed away, its significance as a summary measure of that permits predictions of response time in practical situations remains important. The data base from which to select parameter values is large and useful.

Signal Detection Theory

References

Green, D. M. and Swets, J. A. Signal Detection Theory and Psychophysics. New York: Wiley, 1966.

Mackworth, J. F. Vigilance and Attention. Baltimore: Penquin, 1970.

Swets, J. A. (ed.) Signal Detection and Recognition by Human Observer. New York: Wiley, 1964.

Wickelgren, W. A. and Norman, D. A. "Strength models and serial position in short-term recognition memory." J. Math. Psych., 3, 1966, 316-347.

Description

Signal detection theory is an elegant model of human discrimination performance that provides independent measures of an individual's sensitivity (discrimination acuity) and judgment bias. Sensitivity is effected by factors such as the physical characteristics of the signal to be detected and the "noise" environment in which the signal is embedded, the state of the observer's sensory system, and the nature of the detection task. Judgment bias reflects the effects upon the decision process of various psychological factors, such as signal expectancy, motivation, consequences of the possible signal-response outcomes ("payoff matrix"), and response bias.

The simplest form of the model is applicable to a detection situation in which a signal ("signal plus noise") is presented on some proportion of a series of the trials, randomly selected; on the other trials, the signal is absent ("noise" alone). The observer reports on each trial whether or not he/she detects the signal. From these reports, two independent statistics may be obtained: (1) the proportion of detections when the signal was present, called "hits," and (2) the proportion of detections when the signal, in fact, was absent, called "false alarms." The basic premise of the model is that the internal sensory observation, X, made by a subject on a given trial is a random variable. The distribution of possible values, usually assumed to be Gaussian, arises from random variation in the background "noise," of either external or internal origin. The distribution of possible sensory values when the signal is presented, f_{SN}, is assumed to have a greater mean (and, in some cases, a

the decision criterion. That area measure the probability of reporting a detection (an observation to the right of the criterion) given that a signal was presented (the f_{SN} distribution). Similarly, the predicted probability of a false alarm is given by the area under the f_N distribution which lies to the right of the decision criterion.

Examination of the model makes clear that there is a predicted trade-off between hits and false alarms for a fixed sensitivity. If the decision criterion is increased, the probability of a false alarm is decreased, but so is the probability of a hit. Similarly, if the criterion is relaxed to increase the probability of a hit, the probability of a false alarm also increases. Observer performance is most often presented in a plot of the proportion of hits versus the proportion of false alarms. For a fixed sensitivity, the locus of the data points swept out in this plot with changing β is a curve (concave downward) which rises from the lower left-hand corner for a very strict criterion to the upper right-hand corner for a very lax criterion. This curve is referred to as a "receiver operating characteristic," or an "isobias" curve, and reflects change only in the observer's bias. A family of such ROC curves is generated as the sensitivity of the observer varies. The major diagonal reflects chance performance since hit and false alarm probabilities are everywhere equal. Increasing sensitivity bows the ROC curve more and more towards the upper left-hand corner, which represents perfect performance (a hit probability of 1.0 and a false alarm probability of 0.0).

Model Validation

The basic detection model has been successfully applied to human psychophysical data in various sensory modalities: vision, audition, taste, smell, touch. Moreover, the model has been found useful in describing discriminative performance along different types of sensory dimensions: changes in intensity, changes in quality, changes in duration or rate. (See Green and Swets (1966), and Swets (1964) for a number of such applications). Various extensions of the model have been derived to allow application of the model to tasks other than the simple Yes/No detection task. Forced-choice procedures, confidence rating tasks, and recognition tasks are samples of other tasks to which the theory has been extended. With the addition of axioms regarding learning and forgetting (decay), the concepts of signal detection theory have been used as the basis of a model of human recognition memory (Wickelgren and Norman (1966)).

greater variance) than the distribution, f_N, arising from noise alone trials. For any given observation value, x, we may determine the likelihood (the probability) that a signal was presented, $f_{SN}(x)$, and the likelihood that noise alone was presented, $f_N(x)$.

The model assumes that the observer calculates the likelihood ratio for the observation,

$$\ell(x) = f_{SN}(x)/f_N(x) \ .$$

If the likelihood ratio is greater than some criterion value, C, he/she reports no signal. An alternative description of the decision procedure, equivalent to the likelihood ratio criterion under most conditions, is more intuitive: the observer chooses a criterion value, x_C, on the observation continuum which partitions the dimension into two regions. If the observation value, x, is greater than x_C, then the observer reports signal detection; if less than x_C, he/she reports no detection.

Input Parameters

There are two parameters: d', which specifies the separation of the signal and noise distributions and, consequently, the observer's sensitivity, and β, which specifies the location of the decision criterion and, consequently, the observer's decision bias.

The sensitivity parameter, d', is a measure of the distance between the means of the signal and noise distributions, relative to their common standard deviation. Thus, large values of d' indicate very little overlap of the two distributions, with resulting high sensitivity, whereas, small values of d' indicate a large amount of overlap, with resulting low sensitivity.

The value of the bias parameter, β, determines the criterial observation required for the observer to report a signal detection. A large value specifies a strict criterion such that fewer signal detections will be reported; a low value specifies a lax criterion such that many signal detections will be reported.

Model Output

For given values of the parameters d' and β, the model predicts the probability of a "hit" and the probability of a "false alarm." The predicted probability of a hit is simply the area under the f_{SN} distribution which lies to the right of

Comments

Signal detection theory is useful both as a means of describing the sensitivity and bias of an observer in a particular task, and as a means of predicting how the observer's performance (hit and false alarm rates) will change as either physical or psychological attributes of the task are modified. While the model was initially developed to apply to discrete trial tasks, there has been work on extending the theory to continuous time tasks (vigilance tasks) in which an unspecified number of signals may occur at random points in time over some extended time interval (see Mackworth (1970) and articles by J. P. Egan et al in Swets (1964).

Vickers' Accumulator Model of Discrimination

References

Vickers, D. Evidence for an Accumulator Model of Psychophysical Discrimination, Ergonomics, 13, 1970, p. 37-58.

Description of Model

Building on considerable previous work on modelling discrimination accuracy and latency, Vickers has assembled the most comprehensive model of discrimination performance available. It is representative of a larger class of more special purpose models and conceptual frameworks that formulate choice latencies as a statistical decision process in which the observer samples information from the stimulus display and accumulates evidence over time that increases his/her confidence in choosing a particular response. When the evidence favoring one of the alternatives exceeds a critical level of confidence, c, in favor of one response then that response is executed.

More precisely, the model has been formulated for a two-alternative discrimination process in which an observer makes comparisons between the noisy excitation generated by stimulus A having a mean $\bar{A}$ and variance σ_a^2 and stimulus B having excitation mean $\bar{B}$ and variance σ_b^2. The distribution of differences is converted to standard score form

$$z = (\bar{A}-\bar{B})/\sqrt{(\sigma_a^2 + \sigma_b^2)}$$

and partitioned by use of cumulative normal statistics into the probability p that A > B and the probability q that B > A (where q=1-p).

The representation then departs from the usual statistical decision formulation by postulating that the observer samples a difference magnitude associated with each sample and that these magnitudes are also normally distributed random variables, having mean $\bar{m}$ and $\bar{n}$ for positive (A>B) and negative (B>A) differences. Thus with probability p the observer samples a positive difference having magnitude m.

The sampling process is considered to be periodic and serial, each sample taking a constant time, λ. Thus the mean response latency is given by

$$\bar{L} = \lambda \bar{N} + t_o$$

where t_o is a constant added to account for relatively fixed delays in apparatus and human response systems and $\bar{N}$ is the mean number of samples required to exceed the response criterion C_a or C_b

The author argues that

$$\bar{N} = \frac{r_a}{p} I_p (r_{a+1}, r_b) + \frac{r_b}{q} I_q (r_{b+1}, r_a)$$

where $r_a = \frac{C_a}{\bar{m}}$, $r_b = \frac{C_b}{\bar{n}}$

and $I_p (r_a, r_b)$ is the incomplete beta-function.

Input parameters

To generate relative latency distributions requires estimates of the mean and variance of the stimulus excitation distributions, which may be expressed in terms of the detection parameter d' or as means and variances directly. These parameters, in principle, should be observer independent, but probably are not. The parameters, λ and t_o must be estimated to calibrate the theoretical distributions to empirical latency distributions. Finally, one must estimate C_a and C_b, the values of N at which the sampling process is to terminate. Usually these would be chosen symmetrically and might well be set a priori if enough experience with their use had been accumulated.

Model Outputs

The basic output of the model is the distribution of response time as a function of the response probability given that signal A or B occurred. This amounts to the joint function of latency and probability of a correct response.

Model Validation

The author has shown that his model fits a variety of sources of laboratory data better than alternative statistical decision formulations and in fact resolves some unexplained discrepancies between previous accumulator models and random walk models. To our knowledge it has never been tested in a practical context.

Comments

This model is typical of those found in the research literature. It is more highly developed than many; however, the mathematical formulation deals only with the two-alternative case. It is likely that it could be generalized, at least in a computer simulation version, to multiple alternatives, but then new validation studies would be needed. The principles of psychophysical discrimination underlie many decision and judgment tasks in application environments and in that sense represent an important basic function to be modelled.

Rumelhart's Model of Visual Display Perception

References

Norman, D. A., and Rumelhart, D. E. A system for perception and memory. In D. A. Norman (ed.), Models of Human Memory New York: Academic Press, 1970, 19-64.

Rumelhart, D. E. A multicomponent theory of the perception of briefly exposed visual displays. J.Math.Psych., 1970, 7, 191-218.

Rumelhart, D. E., and Siple, P. Process of recognizing tochistoscopically presented words. Psychol. Rev., 1974, 81, 99-118.

Description

Rumelhart's model describes the time course for the stochastic extraction of feature components from a briefly-presented visual display. The model is capable of predicting the probability of recognizing elements (e.g. alphanumeric characters) in a display under given conditions of display size and quality, and of predicting the latency of correct and incorrect responses. The major assumptions of the basic model (Rumelhart, 1970) can be stated as follows:

(1) When a display of N elements is exposed for T milliseconds, the display is registered in a visual sensory store. Following offset of the display, information in the sensory store decays exponentially, with time constant μ.

(2) Each element in the display is assumed to be composed of a set of physical features. A later version of the model (Rumelhart and Siple, 1974) further specifies that each element (say, an alphanumeric character) is composed of a number of components, where each component is a straight line segment, or a portion of such a segment, which is specific with regard to orientation and retinal location. The number of components making up a straight line segment is assumed to be proportional to the length of the segment.

(3) Components are extracted randomly from the display according to a nonhomogeneous Poisson process, with the rate of extraction, $v(t)$, being equal to a constant v while the display is exposed $(0 \leq t \leq T)$, and proportional to the decaying clarity of the information thereafter:

$$v(t) = \begin{cases} 0 & t<0 \\ \nu & 0 \le t \le T \\ \alpha \nu \varepsilon^{-(t-T)/\mu} & T<t \end{cases} ,$$

where α is a parameter ($0 \le \alpha \le 1$) determined by characteristics of the post-exposure field. (For a dark post-exposure field, α is assumed equal to one.)

(4) The probability that an extracted component belongs to element i of the N elements in the display is given by

$$\sigma_i = \frac{\omega_i}{\sum_{i=1}^{N} \omega_i}$$

where ω_i is the weight corresponding to the attention assigned to position i of the display. In most applications, equal weighting among the elements has been assumed, so that

$$\sigma_i = \frac{1}{N} ,$$

for all i.

(5) In the early version of the model, Rumelhart (1970) assumed that the extraction of C components (a parameter) of a given element was sufficient for recognition of the element. It may be supposed that the magnitude of C will depend upon the physical characteristics and size of the set of possible elements to be recognized. From this, and the above, assumptions, one can calculate the probability that a given element will be recognized at time t (or that it will ever be recognized for a given exposure duration).

In a later version of the model applied to word recognition, Rumelhart & Siple (1974) assumed that, following exposure of a letter string to be identified, the subject reduced the set of all possible response strings to only those consistent with the set of components extracted, and which had no more than C components missing. A response string was reported using a Bayesian rule which weighted the possible responses according to <u>a priori</u> expectations based on word frequency, letter transition probabilities, and other experiential factors.

Input Parameters

The basic parameters of the model are as follows:

μ	$\equiv$	the time constant of the decay of information in the visual sensory store.
ν	$\equiv$	the rate of extraction of components from the sensory store during display presentation.
α	$\equiv$	a parameter describing the degradation of the information in the visual sensory store by the post-exposure field.
$\{\omega_i\}$	$\equiv$	the set of weights describing the attentional bias towards the position of elements in the display (uniform weighting often assumed).
C	$\equiv$	the critical number of components required for recognition of an element.

Model Output

With the parameter values specified, the model predicts the probability that any given element in the stimulus display will be recognized by time t, or that it will ever be recognized (determined by *the limit as* $t \to \infty$). Although not stressed in Rumelhart's development, the model is also capable of predicting mean response time for correct and incorrect responses in recognition tasks.

Model Validation

The model has been applied to a number of tasks in which strings or arrays of alphanumeric characters are presented tachistoscopically to subjects: (1) It accurately predicts Sperling's results on the number of elements recalled correctly as a function of display size, both for the complete and cued partial report conditions. (2) The model also makes accurate predictions of the results of experiments by Estes and his co-workers in which subjects reported which of two "critical" letters was contained in a briefly presented display. (3) It predicts the probability of an error in the recognition of a target character when the stimulus display is followed by a mask after some delay. Finally, (4) Rumelhart has shown that the model can account reasonably for the effects of word (or string) frequency, letter predictability, and letter confusability on the probability of recognizing the letters in briefly presented letter strings.

Atkinson and Shiffrin Model of Human Memory

References

Atkinson, R. C., Brelsford, J. W., Jr., and Shiffrin, R. M. "Multi-process Models for Memory with Applications to a Continuous Presentation Task." J. Math. Psych., 4, 277-300, 1967.

Atkinson, R. C., and Shiffrin, R. M. "Human Memory: A Proposed System and Its Control Processes." In, K. W. Spence and J. T. Spence (eds.), The Psychology of Learning and Motivation, Vol. 2, New York: Academic Press 1968, 89-195.

Description

Atkinson and Shiffrin have proposed a model of human memory which distinguishes permanent memory structures from modifiable, task dependent control processes. The structural components of the system consist of multiple sensory registers, an auditory-verbal-linguistic short-term store (STS), and a long-term store (LTS). Incoming sensory information is automatically registered in the appropriate sensory register, from which it decays rapidly (e.g., within several hundred milliseconds in the visual sensory register). Under the control of modifiable attentional and scanning processes, selected information can be transferred from the sensory register to the STS, the active working memory. The STS is characterized by a limited information capacity, estimated to be on the order of 4-5 items, based on several of Atkinson and Shiffrin's experiments. An item of information is subject to decay from STS in about 30 seconds in the absence of a rehearsal strategy. Rehearsal is an active process in which items are retained in a buffer of limited size and systematically renewed. New items can displace old items in the rehearsal buffer, under voluntary control. The LTS is the permanent memory structure in the system. The "strength" of the representation of an item in the LTS is hypothesized to increase linearly with time while the item resides in the STS. Thus, if an item remains in the STS for an interval t, its strength in LTS is given by σt, where σ is the transfer rate. Once an item leaves the STS, either through decay or displacement, the strength of the LTS trace decays geometrically; if an item has strength I at time o, it will have strength τI at time 1, $\tau^2 I$ at time 2.... The probability of correctly retrieving an item from LTS is assumed to be related to the strengh, I, of the item in LTS by

$$\rho = 1 - (1-g)\epsilon^{-I},$$

where g is the probability of guessing the item correctly. It is further assumed that an item currently residing in the STS will be correctly retrieved with probability 1. Thus, the overall probability of retrieving an item correctly is given by the probability that the item is currently in STS plus the product of the probability of retrieving it from LTS with the probability that it is not in STS.

Input Parameters

There are two parameters which describe the transfer of information into, and out of, the LTS:

σ = the rate of transfer of information about an item from STS into LTS.

τ = the proportion of information retained in LTS per unit time ($1-\tau$ = proportion of information lost per unit time).

If the rehearsal buffer is hypothesized as part of a person's strategy for a given task, there are up to three other parameters which may be specified:

γ = size of the rehearsal buffer

α = probability that a new item (not currently in the buffer) will enter the buffer, displacing an old item.

δ = parameter describing bias towards displacing oldest items in buffer before newest items.

Model Output

This theory asserts that human memory performance is determined not only by the structural components of the system, but also by the control processes which are altered as required by the demands of the task. Thus, a detailed and complete specification of the model requires assumptions about the control processes likely to be used in a given task. Given these further assumptions, the model is capable of predicting memory performance measures such as the probability of a correct response given the number of intervening items presented, the types of items presented, serial position of an item in a list, rate of presentation of items, etc. The model does not address issues related to the time required for storage or recall of memorized information.

Model Validation

The model has been applied to a number of the common experimental verbal learning and verbal memory tasks, with considerable success. These tasks have included several types of paired-associate tasks which stress short-term memory load, and several types of serial learning and free-verbal-recall tasks. There is, unfortunately, some evidence from the studies that the parameter values are task dependent and, perhaps, also subject specific. This fact restricts the generality of the model.

Comments

The strength of this model is its specification of the structure of a rehearsal buffer and its related control processes as a primary strategy in the use of short-term memory. The model stresses the role that may be played by the long-term store in what would otherwise be considered purely short-term memory tasks. On the other hand, the model does not specify either the structure or the organization of information in the long-term store. Thus, the model is primarily useful as a description of human performance in remembering and retrieving sequences of items in real time, where short-term memory and rehearsal processes are important and coding and organization processes are not important.

Movement Index of Difficulty: Fitts' Law

References

Fitts, P. M., "The information capacity of the human motor system in controlling the amplitude of movement," Journal of Experimental Psychology, Vol. 47, 1954, pp. 381-391.

Fitts, P. M., J. R. Peterson, "Information capacity of discrete motor responses," Journal of Experimental Psychology, Vol. 67, 1964, pp. 103-112.

Welford, A. T., "Movement," Chapter V in Fundamentals of Skill, London, Methuen & Co., Ltd., 1968, pp. 137-160.

Description of Model

The model predicts the time required to make a speeded, simple positional movement given the distance to be moved and the accuracy constraints on the movement. It is basically an empirically derived law, although the form of the model may be derived from information theory principles or feedback control principles.

Fitts' basic equation is

$$MT = a + b \log_{10} \frac{2A}{W}$$

where MT is the required movement time; A is the distance to be moved to the center of the target area; W is the width of the target area; and a and b are empirically derived constants.

The model predicts movement time directly, and deals with accuracy implicitly through the definition of W. It has been shown that the basic relationship is valid even when speed of movement is manipulated, as long as effective target width is calculated using the "Crossman Correction" (Welford, p. 147). This correction assumes a Gaussian distribution of hits, and defines the effective target width as that which includes 96 percent of the hits. In most applications, however, the target size is predefined and movement time is the dependent variable.

While this model is very robust, accounting for 90 to 95 percent of the variance in movement times under simple movement conditions, revisions of the equation have been proposed by

Welford to achieve greater precision in a variety of conditions. His modification is given by

$$MT = b \log_{10} \frac{A}{B} + (b-a) \log_{10} C$$

where A, W, a, and b are defined as before and C is interpreted as a minimal scatter of target hits at a point target.

Input Parameters

Movement distance (A) and target width (W) are required input parameters. The constants a, b, and C are assumed to be relatively invariant. Welford assigns them values of .104 sec/bit, .179 sec/bit, and 31 mm., respectively. In Fitts' formulation, typical values for discrete movements are a = -.570 sec and b = .074 sec/bit.

While these parameters appear to be relatively invariant, they should be checked empirically if the application is to other than simple positional movements of a finger or stylus.

Model Outputs

In the form shown, the dependent variable is movement time. As discussed above, however, the model could also be used to predict movement accuracy as a function of speed.

Model Validation

The Fitts' law methodology has been applied in a large variety of settings, ranging from the control of hand movements under microscopic magnification to positioning of a light spot directed by head movements in an application to a photocell-operated typewriter for paraplegics. There is even some indication that the basic equation is applicable to foot movements, but with different parameter values. It also has been shown to work for positioning pegs into holes and when significant weight is attached to the hand. To a first approximation, it will work for fore-aft movements as well as side-to-side movements, although most available data were obtained for the latter case. It does not predict accurately when the ratio of movement distance to target size is less than about 2 to 1.

Comments

This model is potentially applicable to any task or subtask in which the layout of positions on a work place is relevant to the design variables under study and where movement time will be a significant proportion of the total time involved in a particular task.

Intuitive Statistics

NOTE: Like the work on heuristics summarized in Abstract No. 29, laboratory efforts to evaluate man as an intuitive statistician and to compare his performance with that prescribed by normative statistical theory have potential relevance to descriptive modelling of behavior in operational settings. Below we present a summary of the classic paper by Peterson and Beach that highlights those aspects of human performance that are thought to be of importance and that, with some further research, could yield to quantitative expression.

References

Peterson, C. R. and Beach, L. R., "Man as an intuitive statistician," Psychol. Bull., 1967, 68(1), 29-46.

Pitz, G. F., "Sample size, likelihood and confidence in a decision," Psychonomic Science, 1967, 8 (6), 257-258.

Description

The long range goal of the work reported is the development of a theory of human behavior in uncertain environments. Of prime interest are those environments that can be characterized statistically and those tasks for which a model of ideal performance can be specified. For our purposes, the cited paper provides a convenient catalog of observed performance in two major areas:

1) estimates of the characteristics of samples of data
2) the use of samples to infer characteristics of parent populations

The first of these is referred to by the authors as "Intuitive Descriptive Statistics"; the second as "Intuitive Inferential Statistics."

I. Intuitive Descriptive Statistics

1. Judgments of proportion: the relation between mean estimates and sample proportions is described well by an identity function with a maximum deviation of the mean estimate from the sample proportion of 0.3 - 0.5 and an average deviation very close to zero. There may be a very slight overestimation of low proportions and an equally slight underestimate of high ones, depending upon experimental conditions.

2. Judgments of means and variances: subjects familiar with the concept of a mean of a distribution give highly accurate estimates, though there is some variance as a function of sample variance, sample size and rate of presentation of data. There are systematic discrepancies between intuitive judgments of sample variance and actual variance. In general, as means increase, estimates of variance decrease.

II. Intuitive Inferential Statistics

1. Inferences about proportions: in experimental tasks requiring sequential revision of subjective probabilities, it is typically found that subjects are conservative in their behavior, revising their estimates in the appropriate direction as additional data are acquired, but at a rate considerably less than is justified.

2. Inferences about means: inferences concerning the central tendencies of symmetrical distributions are generally accurate. In J-shaped distributions, inferences concerning median and mode are accurate but those relating to the mean are biased toward the median.

3. Inferences about correlations: evidence suggests that subjects do not pay attention to all cells of a 2 x 2 contingency table when inferring correlation. That cell in which favorable outcomes appear together or those associated with the positive diagonal seem to receive primary consideration. As the tabular dimensions increase beyond 2 x 2, inferences become more nearly like those assessed by normative correlation statistics.

4. Determining Sample Size: in experiments in which a subject must infer which of two populations is being sampled, he has continuous control over the number of samples taken, a cost is assigned to each sampled datum, and the dependent variable is the number of data purchased prior to making a decision, it is typically found that although performance is influenced in the proper direction by cost/payoff structure, the magnitude of the influence of this structure is less than that prescribed by ideal models. When the ultimate size of the sample is decided ahead of time by the subject, the cost/payoff structure appears to have no effect on behavior.

When prior probabilities are reduced by defining more alternatives, subjects appropriately purchase larger samples of data. When only two alternatives are involved and prior probabilities are made more disparate, fewer data are purchased. With any number of alternatives or distributions of priors, however, purchasing performance is conservative relative that dictated by the optimal (Bayesian) rule.

Input Parameters

Not applicable

Model Outputs

Not applicable

Model Validation

All of those aspects of performance summarized above were observed under laboratory conditions and most all have been validated in both earlier and later studies using a wide variety of stimulus materials and response modes. Efforts to define parameters in information purchasing tasks within the fixed and optional stopping paradigms and to construct descriptive models have been particularly intense (see, for example, Pitz, 1968, 1969 and Pitz, et al, 1967, 1969).

Comments

In our judgment, observations such as those above are useful in model building activity in two different ways: (1) collectively, they help to establish the forms of relationships between independent variables and performance in situations requiring quantitative inference; and (2) they help to define the ranges over which these relationships can be expected to obtain. Because of the frequency with which the phenomenon of conservatism in sequential sampling has been demonstrated, we believe there is sufficient justification for attempting to capture its basic form in descriptive models of human performance. For any given situation, of course, some prior empirical work must be accomplished to define its range.

With respect to the other indicators of sub-optimal performance, we feel less certain. Perhaps only additional research within the precise contexts of given operational tasks can verify their existence and significance.

Kahneman and Tversky's Heuristics

Note: Recently, psychologists have become interested in discovery and description of the sorts of heuristics, algorithms and rules-of-thumb people employ in the solution of problems requiring inference and reasoning. At this time, some of the most interesting findings can only be described in qualitative terms. This summary contains brief descriptions of two heuristics which are, in our judgment, of potentially great relevanace to quantitative modelling of human performance.

References

Kahneman, D. and Tversky, A. "Subjective probability: A judgment of representativeness." Cog. Psychol., 3, 430-454, 1972.

Kahneman, D. and Tversky, A. "On the psychology of prediction." Psychol. Rev., 80, 237-251, 1973.

Tversky, A. and Kahneman, D. "Belief in the law of small numbers." Psychol. Bull., 76,105-110, 1971.

Tversky, A. and Kahneman, D. "Judgment under uncertainty: Heuristics and biases." Science, 185, 1124-1131, 1974.

The Representativeness Hypothesis

According to this hypothesis, people select or order outcomes on the basis of the degrees to which these outcomes represent the essential features of the evidence they have examined. In doing so, they ignore the prior probabilities of those outcomes and/or the degree of unreliability inherent in the evidence. Where the representativeness of the outsomes is, in fact, matched to the likelihoods, this behavior leads to a statistically appropriate selection (or ordering). In situations where they are not matched, however, the selection (ordering) will be inappropriate.

An important prediction from this hypothesis is that whereas the processes of evaluating a set of data and of predicting an outcome on the basis of the data are independent and should entail differing amounts of uncertainty when the data are less-than-perfectly reliable, people will perform as if the processes and uncertainties are similar. The essential difference between evaluating an input and predicting an outcome is highlighted in an example provided by the authors:

> "Suppose one is told that a college freshman has been described by a counsellor as intelligent, self-confident, well-read, hardworking, and inquisitive. Consider two types of questions that might be asked about this description: (a) Evaluation: How does this description impress you with respect to academic ability? What percentage of descriptions of freshmen do you believe would impress you more? and (b) Prediction: What is your estimate of the grade point average that this student will obtain? What is the percentage of freshmen who obtain a higher grade point average?"

The authors point out that there is surely greater uncertainty concerning the second question than the first, hence, that the prediction should be closer to fifty percent than the evaluation. They predict, however, (and later demonstrate) that subjects will produce approximately the same estimate for both questions since they will consider the best prediction to be that score which is most representative of the input data.

A second prediction resulting from the representativeness hypothesis is that the degree of confidence one has in his prediction from a set of data reflects the degree to which the outcome he has selected is more representative of the set than are other possible outcomes. Thus, one might have more confidence in the prediction of an overall B average in future courses on the basis of B grades in two separate introductory courses than he would have given an A and a C in those courses. This, the authors note, is incompatible with the common multivariate model of prediction in which predictive accuracy is independent of within-profile variability.

The tendency, if it exists, to predict solely on the basis of the input data and to ignore the prior probabilities of chosen outcomes may also be at the root of commonly observed failures to appreciate regression phenomena. In the authors' view, the regression effect typically violates the intuition that a predicted outcome should be maximally representative of the input information.

The Availability Heuristic

With this heuristic, a person judges the frequency or probability of an event or class of events on the basis of the ease with which relevant instances are brought to mind. An example provided by the authors is that of assessing the risk of heart attack among middle-aged people by recalling such occurrences among acquaintances.

As in the case of inference via representativeness, the predictions one is led to under this heuristic are frequently correct, since instances of frequent events are typically recalled better and more quickly than instances of less frequent events. Associated with its use, however, are certain biases which occasionally produce inappropriate judgments: some of these are as follows:

(1) Biases due to retrievability of instances. A class whose distances are easily recalled will, using this heuristic, appear more frequent than a class of equal frequency whose instances are less easily recalled.

(2) Biases due to effectiveness of search set. In tasks requiring estimates of the relative frequencies of different words, the availability heuristic will lead to the judgment that abstract words (e.g., love) are more numerous than concrete words (e.g., story).

(3) Biases of imaginability. In instances in which one must assess the frequency of class whose instances are not stored in memory one may identify a rule or algorithm that can, in turn, generate instances for consideration. The ease with which instances can be generated then become the criterion for frequency assessment. Depending on the characteristics of the rule and the prior likelihoods of the instances generated, the assessment may or may not be accurate.

(4) Illusory correlation. This term refers to a tendency to the frequency of a co-occurence of events which are naturally associated with each other.

Input Parameters

Not applicable.

Model Outputs

We have outlined a number of the empirical expectations that follow from the hypotheses that representativeness and availability heuristics frequently underlie the evaluation of data and the prediction of outcomes. These and others of perhaps lesser importance in the current context can be summarized as follows:

(1) Judgments of the probabilities of future events will tend to be representative of the input data and not to reflect the prior probabilities associated with those events.

(2) Judgments as the probability of obtaining a particular result in a sample drawn from a specified population will be relatively insensitive to the size of the sample.

(3) The confidence one has in a prediction will depend almost completely on the degree to which that prediction is representative of the input data and will be insensitive to the reliability of the data.

(4) The normal tendency for the performance of people (or machines) to oscillate about some mean level - hence, for a better-than-average or poorer-than-average performance to be followed immediately by an "average" performance will fail to be recognized. One possible outcome of this failure may be that punishment will be perceived to be more effective than reward.

(5) When judging the relative size of classes by the availability of their instances, a class whose instances are easily recalled will appear larger than one whose instances are less easily recalled.

(6) If the frequency of words is judged by availability of contexts, abstract words will be judged to be more numerous than concrete words.

(7) The frequencies of co-occurrences of natural associates tend to be overestimated.

(8) The frequency of conjunctive events will tend to be overestimated; the frequency of disjunctive events will tend to be underestimated.

Model Validation

The representativeness and availability hypotheses were formulated on the basis of empirical studies in a wide variety of areas including probabilistic inference, evaluation of student test profiles, recall of material stored in memory, etc. The specific predictions (discussed under "Model Outputs" above) were later examined in carefully controlled laboratory studies by the authors. Those identified have been found to be valid over a wide range of evaluation and prediction contexts, and some (e.g., the overestimation of events having a natural association) seem remarkably resistant to contradictory data.

Comments

As noted in the preface to this summary, we feel that findings of this type have potentially great relevance to the descriptive modelling of human performance. They would appear to

be applicable to many situations in which an operator must predict the future state of a system on the basis of noisy input data, or where he is required to perform fault isolation and trouble shooting tasks. We expect that for such situations, the basic forms of the relationships between input data and evaluation and prediction performance would be relatively easy to determine empirically, though values suitable for precise estimation in either analytic or simulation models might require a relatively extensive program of parametric research.

The point should also be made that since the behavioral phenomena of interest relate largely to man's capabilities as an intuitive statistician, there exists, in the form of prediction from normative statistical theory, a ready standard against which to compare to his performance. The ability to achieve this comparison can aid directly in assessing the relative merits of alternative system aids to the evaluation and prediction performance of the human controller.

The Bayesian Model of Inference

References

Edwards, W., "Behavioral Decision Theory," Ann. Rev. Psychol., 1961, 12, 473-493.

Nickerson, R. S. and Feehrer, C. E. Decision making and training: A review of theoretical and empirical studies of decision making and their implications for the training of decision makers. NAVTRAEQUIPCEN 73-C-0128-1, July 1975.

Description

Given a set of mutually exclusive and exhaustive hypotheses, (H_i) and datum, D, Bayes' rule expresses the probability that a given hypothesis, H_j, is true as a function of $p(D|H_j)$, the probability that D will be observed given that H_j is true, and $p(H_j)$, the probability that H_j is true prior to the observation of D.

$$p(H_j|D) = \frac{p(D|H_j)p(H_j)}{p(D)}$$

where $p(D) = \sum_{i=1}^{n} p(D|H_i)p(H_i)$ n = total number of hypotheses in (H_i)

When a sequence of observations is made, this formulation is applied recursively and the value of $p(H_i|D)$ that is computed as the result of one observation is employed as the $p(H_i)$ for the next.

Since Bayes' rule specifies an optimal policy for aggregation of data against hypotheses and for the sequential revision of probabilities distributed over alternative hypotheses as data are examined, it has served as an important normative framework with which to evaluate the decision making performance of real decision makers. Further, it has been advanced as a useful

decision aid in the operation of complex man/machine data acquisition and data processing systems, particularly those dealing with intelligence information. Recently, application of the rule has been generalized to situations where the reliability of data may be suspect (see Abstract No. 31).

Input Parameters

Exploitation of Bayes' rule requires that the prior probabilities of each of the possible hypotheses be input and that the set of conditional probabilities [$p(D|H_j)$] be defined. As indicated above, the priors [$p(H_j)$] need only be defined for the first iteration in a sequential revision situation.

Model Outputs

Bayes' rule provides an optimal estimate of the conditional probability $p(H_i|D)$ after the examination of each new datum.

Model Validation

Comparison of man's actual inferential performance against the dictates of Bayes' rule has been one of the leading preoccupations in modern decision making research and theory. These efforts have been directed at finer understanding of one of the earliest findings in the Bayesian decision making area, viz. that humans appear to assign a less extreme probability to the likelihood of a given hypothesis after examining a given datum than is justified on the basis of the prescriptive rule.

Despite the acknowledged productivity of this framework, there are, in our judgment, a number of limitations that must be considered in assessing the generality and utility of the rule to real decision making situations. Our concerns are dealt with at some length in Ref. 2 above and will only be summarized here.

1) The rule concerns only that aspect of decision making related to hypothesis evaluation.

2) It requires uncertainty to be expressed in terms of a mutually exclusive and exhaustive set of hypotheses.

3) New hypotheses cannot be added to an existing set of hypotheses once evaluation has begun.

4) Bayes' rule has nothing with the "truth value" of a given hypothesis, only with its statistical support.

5) Incorrect assignment of prior likelihoods may be devastating when hypotheses are evaluated on the basis of few data.

6) Bayes' rule does not provide a direct criterion for "stopping" in a sequential data acquisition situation. Such a criterion must be adduced to the decision framework from outside.

Bayesian Models of Decision Making With Unreliable Data

Note: The following models are presented because of their potential relevance to modelling of controller judgments as to whether or not "noisy" reports of vehicle position indicate "on-course" or "off-course" behavior. They are presented as a group primarily because of their formal similarity.

References

Dodson, J. D., Simulation system design for a TEAS simulation research facility. Report AFCRL 1112 and PRC 194. Planning Research Corporation, Los Angeles, November 15, 1961.

Funaro, J. F. An empirical analysis of five descriptive models for cascaded inference. Rice University, Department of Psychology Research Report Series, Report No. 74-2, May 1974.

Gettys, C. F. & Wilke, T. A. The application of Bayes's theorem when the true data state is uncertain. Organizational Behavior and Human Performance, 1969, 4, 125-141.

Phillips, L. D. & Edwards, W. Conservation in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-357.

Schum, D. A. & DuCharme, W. M. Comments on the relationship between the impact and the reliability of evidence. Organizational Behavior and Human Performance, 1971, 6, 111-131.

Snapper, K. J. & Fryback, D. G. Inference based on unreliable reports. Journal of Experimental Psychology, 1971, 87, 401-404.

Snapper, K. J. & Peterson, C. R. Information seeking and data diagnosticity. Journal of Experimental Psychology, 1971, 87, 429-433.

Description

Unlike most Bayesian models of human information processing, this family of models explicitly considers that the data on which the decision maker must base his judgment have less than perfect reliability. In the first of the two prescriptive models presented below, the source of error arises during the original observation of an event and the observer/decision maker must estimate whether or not the event actually occurred. In the second, the error arises during the reporting of the event to the decision maker by another observer. In this latter instance, the report is made without qualification and the decision maker must formally discount the impact of the reported event. Both models are of the "cascaded inference" type, requiring that the decision maker first estimate the impact of the datum as though it were perfectly reliable and then, in a second step, alter this nominal impact.

1. Prescriptive Approaches

1.1 The Dodson Model

Dodson (1961) considered the situation in which an observer is not certain which of two mutually exclusive events, D_1 and D_2, has occurred, but may be able to make a probability or certitude judgment on the question. He suggested that in order to calculate the posterior probability of a hypothesis in this case, one should calculate its value, given each of the possible events, and then take a weighted sum of these values, the weights being the probabilities that the observer attaches to the event possibilities. Given only two possible data, the calculation may be represented as follows:

$$\xi(H_i|D) = \psi(D_1)p(H_i|D_1) + \psi(D_2)p(H_i|D_2)$$

where $\xi(H_i|D)$ is the posterior probability of H_i, taking the observer's uncertainty into account, and $\psi(D_i)$ is the probability that the observer attaches to the possibility that he has observed event D_i. More generally, given n possible events and the assumption that the observer can attach a probability to each of them, the formula might be written as:

$$\xi(H_i|D) = \sum_{j=1}^{n} \psi(D_j)p(H_i|D_j).$$

Substituting the Bayesian formula for $p(H_i|\Omega_j)$, yields the expression

$$\xi(H_i|D) = \sum_j \psi(D_j) \frac{p(D_j|H_i)p(H_i)}{\sum_i p(D_j|H_i)p(H_i)} .$$

1.2 The Models of Schum and DuCharme

In this model, there are two possible hypotheses, H_1 and H_2, two possible data events, D_1 and D_2, and two possible reports of the event that occurred, d_1 and d_2. The decision maker's task is to determine $p(H_i|d_j)$. According to Bayes' rule:

$$p(H_i|d_j) = \frac{p(d_j|H_i)p(H_i)}{p(d_j)}$$

which can be manipulated to obtain:

$$p(d_j|H_i) = \sum_k p(D_k|H_i)p(d_j|H_i \Omega D_r)$$

Given this, a likelihood ratio (Λ) adjusted for the diagnosticity of the datum can then be derived.

$$\Lambda = \frac{p(d_j|H_i)}{p(d_j|H_k)}$$

The authors distinguish four decision "cases" which differ with respect to the symmetry of the impact of observing a given event on $p(D|H)$ and $p(d|D)$ and derive the optimal Λ for each. A summary of each of these follows:

<u>Case I</u>: Symmetric $p(D|H)$: Symmetric $p(d|D)$
$p(D_1|H_1) = p(D_2|H_2)$; $p(d_1|D_1) = p(d_2|D_2)$.

$$\Lambda_{s,s} = \frac{pr+(1-p)(1-r)}{(1-p)r+p(1-r)}$$

where $p \equiv p(D_i|H_i)$

$1-p \equiv p(D_j|H_i), \quad j \neq 1$

$r \equiv p(d_i|D_i)$

$1-r \equiv p(d_j|D_i), \quad j \neq 1$

Case II: Asymmetric $p(D|H)$; Symmetric $p(d|D)$

$p(D_1|H_1) \neq (D_2|H_2)$; $p(d_1|D_1) = p(d_2|D_2)$.

$$\Lambda_{a,s} = \frac{p_1 r+(1-p_1)(1-r)}{p_2 r+(1-p_2)(1-r)}$$

or, equivalently,

$$\Lambda_{a,s} = \frac{p_1+k}{p_2+k}$$

where

$$k = \frac{1-r}{2r-1}, \quad r \neq .5.$$

$p_1 \equiv p(D_1|H_1)$

$1-p_1 \equiv p(D_2|H_1)$

$p_2 \equiv p(D_1|H_2)$

$1-p_2 \equiv p(D_2|H_2)$, and r and 1-r are defined as above

Case III: Symmetric $p(D|H)$: Asymmetric $p(d|D)$
$p(D_1|H_1) = p(D_2|H_2)$; $p(d_1|D_1) \neq p(d_2|D_2)$

$$\Lambda_{s,a} = \frac{pr_1+(1-p)(1-r_2)}{(1-p)r_1+p(1-r_2)}$$

or if $p\neq 1$ and $r_2\neq 1$,

$$\Lambda_{s,a} = \frac{c\left[\frac{p}{1-p}\right]+1}{c+\left[\frac{p}{1-p}\right]}$$

where

$$c = \frac{r_1}{1-r_2}.$$

Case IV: Asymmetric $p(D|H)$; Asymmetric $p(d|D)$
$p(D_1|H_1) \neq p(D_2|H_2)$; $p(d_1|D_1) \neq p(d_2|D_2)$.

$$\Lambda_{a,a} = \frac{p_1r_1+(1-p_1)(1-r_2)}{p_2r_1+(1-p_2)(1-r_2)}$$

or if $r_1 \neq (1-r_2)$

$$\Lambda_{a,a} = \frac{p_1+b}{p_2+b}$$

where

$$b = \left[\frac{1-r_2}{r_1-(1-r_2)}\right].$$

and the remaining parameters are defined as above.

2. Empirical Approaches

In addition to an interest in developing prescriptive models against which actual performance can be assessed, investigators

in this area have derived a number of empirical models, the most significant of which are summarized below.

2.1 Snapper and Fryback

These authors suggest that in dealing with unreliable data, decision makers first estimate the likelihood ratio as though the data were completely reliable, then adjust this ratio by multiplying it by the reliability of the quotient, and, finally, apply the adjusted ratio to the calculation of posterior odds.

Stage 1: compute $\tilde{\lambda} = rL$

Stage 2: compute $\tilde{\Omega}_1 = \tilde{\lambda}\Omega_0$.

where r = the reliability of the report
L = the likelihood ratio
$\tilde{\Omega}_1$ = the adjusted posterior odds
Ω_0 = the (unadjusted) prior odds

2.2 Gettys, Kelly and Peterson

This model assumes that the decision maker estimates posterior odds assuming the most likely event is true, and then adjusts the odds to reflect the reliability of the source.

State 1: compute $\Omega_1 = L\Omega_0$

Stage 2: compute $\tilde{\Omega}_1 = r\Omega_1$.

2.3 Edwards and Phillips

These authors present evidence suggesting that posterior odds are described by a single stage process, as follows:

$$\Omega_1 = L^c\Omega_0$$

where c varies with L.

Input Parameters

As with other Bayesian models, initial odds and the conditional probabilities $p(D_i|H_i)$ are required input. In addition, the value of ψ is required in the Dodson model and r is required in all others.

Model Outputs

The Dodson and Schum and DuCharme models provide optimal estimates of diagnostic impact for the cases each considers. The empirical models provide descriptions of actual (non-optimal) performance.

Model Validation

Research indicates that although the behavior of decision makers in choice experiments is clearly influenced by the stated reliability of the reporting source, the performance is not in accord with prescriptive formulations. In general, results suggest that subjects exhibit the classically expected conservatism with data of relatively high reliability and considerably less conservatism than justified with data of low reliability. The collective effort to construct empirical models of the decision process highlights the departure from ideal performance.

Comments

As noted at the beginning of this summary, this family of models may have some application in situations where a decision maker must process and eventually decide the meaning of a set of noisy state data. In the RPV monitoring task, such a situation may occur where the controller must decide whether or not a sequence of position reports, known to have some probability of error, signifies a departure from planned course.

A shortcoming with most experiments in this area is that subjects are not usually apprised of the correct method of adjusting diagnosticity of unreliable reports. Their performance given this information should be of interest in the modelling of skilled behavior.

Dunlay, et al. Models for ATC Conflict Perception

References

Connolly, D. W. and McCosker, W. R. (1970). Human Factors in Use of Terminal Radar (Analogue) Display Systems. U.S. Federal Aviation Administration, National Aviation Facilities Experimental Center, Report No. FAA-RD-70-66.

David, H. (1967). Maneuvers Recommended by Air Traffic Controllers Observing Electronic Simulations and Their Estimated Consequences. Loughborough University of Technology, Department of Ergonomics and Cybernetics.

Dunlay, W. J., Horonjeff, R., and Kanafani, A. "Models for estimating the number of conflicts perceived by air traffic controllers," Univ. of Calif., FAA Contr. No. DOT FA-72-WA-2827, Dec. 1973.

Mangelsdorf, J. E. (1955). Variables Affecting the Accuracy of Collision Judgments on Radar-Type Displays. Ohio State University, Laboratory of Aviation Psychology, Wright Air Development Center, Report No. WADC-TR-55-462.

Description

The goal of Dunlay, et al's work is the development of models of the reactions of human air traffic controllers and of the distributions of aircraft in controlled airspace that could be employed in combination to estimate the number of potential conflicts requiring controller intervention, and, therefore, an estimate of controller workload. The effort appears to be unique in that the authors consider not only the separation at which controllers should take action, as defined by FAA separation standards, but also the distribution of separations at which they have been observed to take action.

Three explicit sources of error are considered in connection with the controller's perception: (1) limitations on the accuracy with which separations between displayed targets can be judged; (2) limitations on the accuracy with which target tracks can be extrapolated when estimating the probability of future intersection with other tracks; and (3) limitations on the accuracy with which the radar targets themselves are displayed. As the authors note, few data exist with respect to the first two of these error sources. The data employed to build elements of the model are those of Connolly and McCasker (1970), Mangelsdorf (1955), and David (1967).

The consequences of these limitations are of interest in two different conflict situations: (1) those which occur when one aircraft overtakes another on the same heading; (2) those which occur when the (projected) tracks of aircraft on different headings intersect. The total perceptual error (ε) is assumed to be the sum of two independent random variables: (1) total extrapolated judgment error (ε_1), and (2) radar system error,

$$\varepsilon = \varepsilon_1 + \varepsilon_2$$

$$\text{where } \varepsilon_1 \sim \left\{ \begin{array}{ll} N(-0.2 \text{ nmi.}, 0.6 \text{ nmi.}) & \text{for overtakes*} \\ N(-0.5 \text{ nmi.}, 2.7 \text{ nmi.}) & \text{for intersections**} \end{array} \right\}$$

$$\varepsilon_2 \sim \left\{ N\ (0.1 \text{ nmi.}, 0.4 \text{ nmi.}) \right\}$$

and the notation N{a,b} is interpreted as a normal distribution with mean a and standard deviation b.

The probability of a conflict given an actual separation, S, is given as

$$P\{\text{Conflict} | S = s\} = P\{(s+\varepsilon)\} \leq S_D\}$$

where S_D is a decision threshold at which the controller must intervene.

Thus, the value of S_D for which the a priori level of performance (say, that a controller be able to detect potential violation of a 5 nmi rule 95% of the time) for overtakes can be expressed as

$$P\{(5+\varepsilon)\} \leq S_D\} = .95 \left[\frac{S_D - (5-0.1)}{0.7} \right] = 6.05 \text{ nmi.}$$

and for crossings,

$$P\{(5+\varepsilon) \leq S_D\} = .95 \left[\frac{S_D - (5-0.4)}{2.7} \right] = 9.04 \text{ nmi.}$$

*Connolly & McCosker data

**David, Mangelsdorf data

with corresponding conditional conflict probabilities,

$$P\{\text{Conflict}|S = s\} = P\{(s+\varepsilon) \leq 6\} = \phi\left[\frac{6-(s-0.1)}{0.7}\right]$$

and

$$P\{\text{Conflict}|S = s\} = P\{(s+\varepsilon) \leq 9\} = \phi\left[\frac{9-(s-0.4)}{2.7}\right]$$

where ϕ is the desired probability of detection (0.95 in this example).

These models are employed in combination with these predicting particular track relationships in a volume of air traffic in order to estimate the expected number of conflicts over specified time periods and, as a result, the expected communication workload of the controller.

Input Parameters

Display resolution, controller extrapolation error as a function of time into the future and of track angle (relative to other aircraft of concern), normative decision threshold and desired level of correct identification are required inputs.

Model Outputs

The primary outputs of the model summarized above are estimates of the distances at which aircraft in potential overtake and crossing situations will be perceived by the controller to be in apparent conflict, as a function of specified levels of desired performance.

Model Validation

As noted earlier, few data exist with respect to the types of controller errors of interest to the author. That controllers sometimes act considerably in advance of the decision time specified in the regulations is, however, a well-known phenomenon. Thus, the basic character of the model's output is valid. Validation studies of the precise values output by the model have, however, not been accomplished.

Comments

The authors have, in our judgment, done a creditable job in using the few data that exist to model a complex phenomenon. The sources of error identified by them and the basic format of the model are intuitively appealing, though attention might also be given to the cost/payoff structure implicit in the situation they are attempting to model.

We believe that the model, if valid, may be of utility in the modelling of RPV controller behavior in those situations where RECON and ELINT vehicles must be recirculated for purposes of accompanying STRIKE vehicles approaching handoff. It might have additional utility in modelling of patching performance, where there is need to extrapolate from a given track the possible point of intersection with a desired track.

Arad's Model of Air Traffic Controller Workload

References

Arad, Bar-Atid, Golden, B. T., Grambard, J. E., Mayfield, C. E., and van Saun, H. R., Control Load, Control Capacity, and Optimal Sector Design, FAA Report RD-64-16, Project No. 102-11R, December 1963.

Arad, Bar-Atid, Notes on the Measurement of Control Load and Sector Design in the Enroute Environment, Appendix to final report on Project 102-11, June 1964.

Ratner, R. S., Williams, J. O., Glaser, M. B., Stuntz, S. E., and Couluris, G. J., The Air Traffic Controller's Contribution to ATC System Capacity in Manual and Automated Environments, FAA Report RD-72-63, Vol. II, June 1972.

Description

Arad's model represents the first major attempt to generate a general model of air traffic controller workload. The original application of the model was to identify optimal sector geometrics to minimize overall workload for a given traffic volume.

The total load, L, imposed on an air traffic controller is divided into several components:

$$L = L_0 + L_1 + L_2 + L_3 + \dots + L_n$$

L_0 is a "background load" component that includes all loads purely internal to the system and unrelated to sector geometry and size, traffic volume, or other measureable phenomena.

L_1 is a "routine load" component associated with the minimum control functions required to move an aircraft into and out of a sector when no interaction with any other aircraft is considered. This component is directly proportional to the number of aircraft in the system per unit time.

L_2 is an "airspace load" component involving expected numbers of conflicts of various types per unit time and the work required to resolve each type. Expected conflicts depend heavily upon traffic flow patterns through a sector, and several subcomponents are analyzed. Generally, this component is proportional to the square of the number of aircraft transiting the sector per unit time.

L_3 through L_n are "induced load" components that arise from coordination requirements with adjacent sectors beyond basic handoff procedures (which are included in L_1).

The underlying structure of Arad's model is straightforward, and most of the load subcomponents are developed from intuitively satisfying models. A large number of constants, parameters, and proportionality constants arise, however, which must be specified before the model can be applied.

To quantify the model parameters, Arad employed several hundred "snapshots" of particular air traffic control situations, which were presented to air traffic controllers in pairs. The controllers were asked to visualize the situations presented and to indicate which of each pair represented the higher workload. The "snapshot" pairs presented had been selected carefully to explore the effects of certain variables and the interactions between them. Statistical analyses of the controllers' choices were used to quantify the parameters.

The structure of Ratner's model (called RECEP, for RElative Capacity Estimating Procedure) is almost identical to that of Arad's model. This model focuses more closely on the times required to deal with the various events that arise. Its parameters were determined by videotaping sequences of actual air traffic control activity, replaying them, and asking the controllers why they took each action they did and when they had decided to take this action. Specific event-processing times were then deduced from a detailed time study of the annotated videotape.

Input Parameters

The geometry of the sector and all airways within it must be specified, along with statistical descriptions of the air traffic within the sector and of potential conflicts that can occur.

Model Output

The models yield scalar representations of overall expected controller workload for each particular situation specified. Workload can also be determined as functions of particular variables, but representations can become complex very rapidly as the number of independent variables is increased.

Model Validation

The judgments of different controllers in Arad's "snapshot" tests were found to be highly consistent. The model thus appears to be internally consistent, and should be capable of accounting for at least some of the discrepancies that have been shown to exist between controllers' perceptions of workload in operational situations and nominal workload measures computed by simpler, traffic-counting procedures. Unfortunately, there is no evidence available in the literature surveyed to indicate that field validation of the model has ever been attempted.

RECEP model parameters were obtained from analyses of field data, and the model has been used to predict the effects of automating certain ATC procedures. Actual validation of the model, however, was limited to a comparison of various decision time estimates obtained in different centers.

Comments

Despite its lack of operational validation, this model may provide a valuable framework on which to build models of controller workload in situations similar to air traffic control.

Sheridan's Model of Process Sampling Frequency

Reference

Sheridan, T. B. "On how often the supervisor should sample," Proc. Inter. Sym. on Man-Machine Systems, Ergonomics Research Society, IEEE, Sept., 1969, V.3.

Description

This model sets forth a prescriptive framework for determining how often a supervisor should sample a process under his control as a function of the extent of his control between samples and the cost of sampling. The model is based upon Bayesian concepts and makes the following assumptions:

1) The functional relationship between the value (V) of the system's independent input (X) and control input (Y) is known and the supervisor will attempt to maximize the expected value.

2) A stationary prior distribution of X exists and is known.

3) After each sample is taken, the parameters of the distribution regress monotonically toward those of the prior distribution.

Given the general formulation of the expected value of a control input,

$$\langle V|Y\rangle = \int_k \langle V|XY\rangle\{X\}$$

where $\int$ is a generalized summation over all X, and {X} is the a priori probability of X. Three specific cases are considered. The first of these is the case where only the a priori distribution {X} of X is known. The second is where the supervisor has perfect knowledge of X (presumably as a result of continuous sampling). The third, and perhaps most interesting, is where, because of the cost of each observation, the supervisor must select a strategy of intermittent sampling/control. An aspect of additional interest is that in the second and third cases, a preposterior analysis can be conducted with respect to X and {X} such that if an optimal Y were to be chosen for each X, $\langle V|Y\rangle$ could be assessed and an optimal supervisory policy adopted prior to X's actual occurrence.

A final step in the development of the model is the addition of cost-of-observation and sampling interval parameters which enable the selection of an optimal sampling interval (or, alternatively, an optimal sampling frequency).

Input Parameters

The prior distribution, $\{X\}$, of X, the expected value relationship, $<V|XY>$, and the cost of an observation must be input.

Model Outputs

The model specifies the optimal time between samples or optimal sampling frequency for the intermittent sampling case. If the case where the a priori distribution is known and the case where perfect information is available are also analyzed, the difference between respective values of $<V|Y>$ can be considered to be the expected value of continuous supervision/control.

Model Validation

Because it is prescriptive in nature, no attempts have been made to validate the model.

Comments

Although not a model of human behavior, this framework may have utility in applications where one wishes to ascertain the impact on system performance of optimal decision behavior under given cost and "knowledge" constraints. Additionally, it appears useful as a reference against which the actual performance of controllers might be assessed. In this latter context, it might also be modified (possibly by the introduction of non-linear value and memory-decay functions) such that it could serve as a descriptive model of human performance.

Drury's Model of Search and Decision Making Processes in Materials Inspection

References

Drury, C. G., "Inspection of sheet materials - model and data," Human Factors, Vol. 17(3), 1975, pp. 257-265.

Drury, C. G., "The effect of speed of working on industrial inspection accuracy," Appl. Ergo., Vol. 4(1), 1973, pp. 2-7.

Sheehan, J. J. and C. G. Drury, "The analysis of industrial inspection," Appl. Ergo., Vol. 2(2), 1971, pp. 74-78.

Description of Model

This model was formulated to account for the empirical finding that as quality control inspection time increases, false rejections of "good" items increase as false acceptances of "faulty" items decrease.

The model assumes that an inspector's response represents the outcome of a two-stage process that initially detects possible flaws, then subsequently compares them against an implicit or explicit criterion in order to reach a decision as to whether candidate items should be accepted or rejected.

On the basis of Crossman's (1955) data on discriminability and his own research on inspection time, the author formulates the following model for the probability of accpeting an item as of time t(s)

$$p(\text{Accept}) = \exp\left(\frac{-t(s)}{t_1(s)}\right) + p(f|a)\left(1-\exp\left(\frac{-t(s)}{t_1(s)}\right)\right)$$

where $t(s)$ is the time taken to locate a flaw; $\bar{t}_1(s)$ is a scale parameter dependent upon search area, contrast, size, etc.; and $p(f|a)$ is the probability of finding a flaw that is, in fact, acceptable.

The probability of rejecting an item as of that time is given as

$$p(\text{Reject}) = p(f|u)\left(1-\exp\left(\frac{-t(s)}{\bar{t}_2(s)}\right)\right)$$

where $p(f|u)$ is the probability of finding a flaw that is, in fact, unacceptable; and $t(s)$ and $t_2(s)$ are defined as above.

Input Parameters

The *a priori* probabilities of acceptably and unacceptably flawed items and the criterion for flaw acceptability are required input parameters, as are the values of the scale parameters, $\bar{t}_1(s)$ and $\bar{t}_2(s)$.

Model Outputs

The model predicts that as search time increases, the probability of accepting a good item will decrease and the probability of rejecting a faulty item will increase. Further, it predicts that the overall inspection per item will be equal to the sum of search time and decision time.

Model Validation

An experiment conducted by the author involving a visual search of flat glass yielded the expected exponential relationship between cumulative probability of locating a flaw and the time taken to locate it.

An experiment on decision making indicated that as the difference between acceptable and unacceptable flaws decreased, decision time and errors increased. Less variation was found in time than in error, however.

Comments

The utility of the model is unknown at this time. However, it would appear that the simplicity of the general framework - that is, the division of the inspection task into a processes of detection and binary classification - should make it applicable to a wide variety of monitoring tasks where a decision rule can be stated.

To the extent that diameter (or area) of a flaw is similar to lateral deviation of a vehicle from an expected path, the model might be used to predict the success with which monitors perform in the context of different display scales and time stresses.

Thomas' Model of Combined Manual and Decision Tasks

References

Thomas, M. U., "A human response model of a combined manual and decision task," IEEE Trans. Sys., Man and Cybernetics, SMC-3, Vol. 5, 1973.

Thomas, M. U., "Some probabilistic aspects of performance times in a combined manual and decision task," Ph.D. dissertation, U. Michigan, Ann Arbor, 1971.

Description of Model

The purpose of this model is to predict the time required to perform a task consisting of n subtasks, in one of which a temporally uncertain signal occurs that requires the subject to choose one from among m alternative responses.

The model considers a repetitive manual task composed of N subtasks, $Q_1 \ldots, Q_n$. One of these subtasks, Q_j, receives "input stimulations" from a stationary random process $I(X)$, resulting in a requirement for a choice of responses from a set, L.

Two input process structures are modelled:

(1) An aperiodic Markov process described by an L x L transition matrix with probabilities

$$P = [p_{ij}]$$

where $p_{ij} = P\{X_k = x_j | X_{i-1} = X_i\}$ $\quad i,j = 1,\ldots,L,$
$\quad k = 1,2,\ldots$

(2) A non-Markov process with discrete probability function

$$p = (p_1, \ldots, p_L)$$

where

$$p_j = P\{X_k = x_j\} = P\{X_k = x_j | X_{k-1} = x_i\}$$

$$i,j = 1, \ldots, L,$$

$$k = 1, 2, \ldots$$

Error rates are assumed to be constant and small, yielding a vector of random variables.

$$T(\tau_k) = (t_1\ (\tau_k), \ldots\ t_N\ (\tau_k))$$

which are the performance times associated with subtasks Q_1, ... Q_N at cycles k=1,2,...

I(X) is related to $T(\tau)$ via the information measure:

$$(1) \quad H(P) = -\sum_i P_i \sum_j P_{ij} \log_2 P_{ij} \quad \text{bits}$$

for the Markov process, and

$$(2) \quad H(p) = -\sum_j p_j \log_2 p_j \quad \text{bits}$$

for the non-Markov process. Drawing on the results of Hick, (1952), the author notes that the linear relationship

$$\hat{\bar{t}}_{j*} = a + bH(p) \qquad a > 0,\ b > 0$$

where $\hat{\bar{t}}_{j-1,j}$ is the decision time between completion of subtask j-1 and subtask j, might be expected to obtain for

$0 \leq H(p) \leq 3.3$ bits when $p_j = 1/L$ and $L = 1, \ldots, 10$ for both Markov and non-Markov I(X).

Input Parameters

The model requires specification of the number of subtasks, Q_j, the number of decision alternatives, and the structure (Markov or non-Markov) of the event generator I(X).

Model Output

The model predicts total performance time of n subtasks as a function of the structure of I(X).

Model Validation

Though it has been employed in a number of simple laboratory experiments, there has been no independent validation of the model.

Comments

The formulation may be useful in modelling tasks requiring the close c-ordination of discrete motor movements in cases where those tasks are performed by skilled subjects. An interesting empirical finding of the study conducted in the context of this model is that once subtasks are fully learned, they can be treated as separate entities for modelling purposes.

Air-To-Ground Acquisition Models

Note: The reference cited below contains a valuable summary of visual detection and recognition models developed in connection with military contracts over the period 1946 to 1972. One or more of these may be useful in modelling the acquisition process during the RPV terminal phase. Our summary of this reference includes only those aspects of two of the formulations that appear to have immediate utility in that phase. The reader is referred to the original article for fuller discussion and for pertinent comments relating to models of this type.

Reference

Greening, Charles P. "Mathematical modelling of air-to-ground target acquisition," Human Factors, 1976, 18(2), 111-148.

Description

Target acquisition models differ in a number of important respects. One is the extent to which they regard the acquisition process as episodic -- that is, consisting of statistically independent tasks such as search, detection, recognition -- or as unitary and unanalyzable. A second difference relates to whether or not a detection lobe concept (cf. Lamar, et al, 1947, 1948) is employed to model the interaction between the geometry of the search area and the visual search process. A third difference has to do with the characteristics of the search itself -- whether or not it is concerned with acquisition of targets which stand in "clear" (geometric, luminance, etc.) distinction with the surroundings -- and with the observer's knowledge about what he is looking for in advance.

1. MARSAM II

Of the models employing a lobe concept, the Multiple Airborne Reconnaissance Sensor Assessment Model (MARSAM II) developed by Honeywell is one of the more complete. It is modular and contains, in addition to submodels of hardware sensor processes, a submodel of the human observer. Elements of the latter are (1) a search function, determined by size of search area, available time, time and glimpse allocation,

line of sight and time varied conditions; (2) a target detection function, dependent upon contrast and size of target luminance and "distinctiveness"; (3) a discrimination function, determined by fixation area and target/non-target density; and (4) a recognition function, determined by acuity threshold, angular rate and critical dimension.

Since the functions are independent, the probability of detection (P_d) and the conditional probability of recognition given detection are:

$$P_d = P_{los} \cdot P^*_{ds} \cdot P^*_{d1} \cdot P^*_{d3}$$

and

$$P_r = P_d \cdot P^*_r$$

, respectively,

where, P_d = probability of detection

P_{los} = probability of the existence of a line of sight to the target

P^*_{ds} = probability of fixations and dwelling upon a target element

P^*_{d1} = probability of detectability

P^*_{d3} = probability of no confusion between target and non-target objects

P^*_r = probability of recognizing a detected target, and

P_r = probability of recognition given detection

Where appropriate, sub-models provide input to terms in this expression. For example, values of target detection parameters are predicted on the basis of an empirical function derived from Blackwell's (1946) laboratory data on visual contrast thresholds, while recognition performance is predicted from the model

$$P_r^* = 1-\exp\left[-k(\beta/\alpha-3.2)^2\right] \quad \text{for } \beta/\alpha \geq 3.2\text{ ; } P_r^* = 0 \text{ otherwise}$$

where P_r^* = probability of recognizing target

β = characteristic angular subtense of target

α = resolution capability of the eye at given apparent contrast

k = a constant

2. Rockwell/Autonetecs Model

Like MARSAM II, this model embodies an episodic view of the acquisition process. Unlike MARSAM II, it does not utilize a visual lobe concept. There are two phases: (1) search and (2) detection and recognition. The search process is not supported by a submodel; instead, a single quantity drawn from empirical research is input. Submodels underlying detection and recognition are, respectively

$$P_D = \exp\left[-\left(\frac{2\alpha^2}{\beta^2(C-C_T)}\right)^m\right] \qquad m = 1 \text{ for } \alpha/2\,/(C-C_T) > 2/\beta^2, \quad m = 2 \text{ otherwise}$$

and

$$P_R = \exp\left[-\left(\frac{8\alpha^2}{\beta^2(C-C_T)}\right)\right]^m \qquad m = 1 \text{ for } \alpha^2/(C-C_T) > 8/\beta^2, \quad m = 2 \text{ otherwise}$$

where P_D = probability of detection

α = resolution capability of eye at C = 1

β = angular subtense of target

C = apparent contrast

C_T = threshold contrast

As can be inferred by comparing these models, the only difference between detection and recognition is the level of resolution required by the geometry of the target.

A cumulative probability of detection/recognition associated with decreasing range-to-target can be determined by summing "single glimpse" probabilities of detection over fixation intervals available from the time the target first becomes visible to the desired minimum range.

$$Pcum = 1 - \prod_i (1 - Psgi)$$

where Pcum = the probability of detection after i independent glimpses

Psgi = the single glimpse probability = $P_i e^{-(r_s/r_o)^2}$, and $P_L e$ is the probability of looking at the target, r_s is the resolution capability of the eye, and r_o is the resolution required to detect or recognize the target

Input Parameters

The brevity of this summary makes somewhat inappropriate a listing of the many inputs required for application of the models. In general, it can be said that information of the following types is necessary:

1. Dimensions of target
2. Angle of visual axis
3. Target/background contrast
4. Visual contrast threshold
5. Target/background luminance
6. Size of visual lobe (MARSAM II and other lobe models only)

7. Size of area to be searched
8. Target/background luminance
9. Available search time
10. Resolving power of eye

Model Outputs

The major outputs of both MARSAM II and the Autonetics/Rockewell models are (1) the probability of detecting a target and (2) of recognizing a target. In addition, the latter model provides an estimate of cumulative probability of detection.

Model Validation

As the author points out, MARSAM II, along with the other episodic models discussed in his paper, develops from a combination of submodels that were, for the most part, individually validated during earlier investigations. The exception to this generalization is the search model, which remains unvalidated.

The Autonetics model has been validated, with good results, against simulator data in an experiment in which recognition performance was predicted as a function of physical range.

Comments

Both of the models summarized above, as well as others reviewed by Greening, appear to provide results of some generality across different types of target detection and recognition situations. As noted earlier, either or both may be useful in the simulation of RPV terminal phase operations. As the author suggests, however, major attention is given to those parameters which are reasonably well understood and rather less attention is given to such things as the character of ground clutter, sophistication of the observer, the differentiation of multiple targets, etc. -- factors which might be of extreme importance in field conditions. Also, it is the case that only the acquisition process itself is of concern and not the problem of vehicle control, terrain avoidance, etc. To be used successfully in the simulation of a total process, they must be teamed with one or more models that predict performance on these other dimensions. Unless these produce estimates that are at least as accurate as those of the target acquisition models appear to be, the precision of the latter and the effort required to specify their input may be largely wasted.

Rouse's Model for Cognitive Prediction of Future States

Reference

Rouse, W. B. "A Model of the Human in a Cognitive Prediction Task," IEEE Trans. on Sys., Man and Cyber., Vol. SMC-3(5), 1973, pp. 473-477.

Rouse, W. B. "Cognitive Sources of Suboptimal Human Prediction," Ph.D. dissertation, M.I.T. Cambridge, Mass., Sept. 1972.

Description

This model was constructed for the purpose of assessing human performance in certain tasks in which events occur on a slow-enough time scale so that limitations in (human response) accuracy are due more significantly to faulty memory and errors in cognition than to limitations in physiological response.

The model considers optimal prediction of future states of discrete linear dynamic systems given by:

$$x_{N+1} = c_o y_{N+1} + C^T X_N$$

where $X_N = \begin{bmatrix} x_{N_1} \\ x_{N_2} \\ x_{N_n} \\ \vdots \end{bmatrix}$ the system state at time N

$C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \end{bmatrix}$ a vector of constants

y_{N+1} = the input at time N+1 sampled from a zero-mean Gaussian process, and

c_o = a constant

For purposes of estimating a future state, it is assumed that the human first identifies the system and gains knowledge of the input characteristics by collecting as data X_N and x_{N+1}, X_{N-1} and x_N, X_{N-2} and x_{N-1}, etc. From these data, he derives an estimate of C which he applies to his prediction.

Critical in the model is the concept of information loss over time. Two aspects of this process are formally represented with negative exponentials: (1) a limitation on the amount of data that can be retained in memory; and (2) the tendency for states that have occurred recently to be remembered better than those that occurred at an earlier time.

A second important component of its model is the concept of observation noise. It is assumed that the standard deviation σ_x of the human's estimates of a quantity x are given by the equation:

$$\frac{\sigma_x}{|x|} = F \quad , \quad \text{a constant}$$

Input Parameters

In addition to system dynamics, the model requires input of the memory function (number of past events) and σ_x.

Model Outputs

The model provides an estimate of the human's expectation of C as a function of memory length and observational accuracy. Within the definition of the model, this represents an optimal output which can be compared against the performance of real subjects.

Model Validation

The author reports an experiment in which subjects were required to predict the horizontal displacement of the last member of a set of eleven dots displayed on a CRT as a function of various system dynamics and input variances. Subjects both familiar and unfamiliar with the independent variables were employed. Outcomes accorded well with predictions of the model. It was also found that changes in the memory parameter of the model had much more influence on the goodness-of-fit with experimental data than did changes in the observational noise parameter.

Comments

The real utility of this model is unknown at this time. Its primary value may be that it provides a framework for assessing performance in systems characterized by slow, discrete changes. However, as the author notes, its successful use requires careful characterization of the time scale, since others (e.g., Kleinman et al, 1971) have found physiological limitations to be of critical importance in tasks requiring rapid response, while still others (e.g., Rapaport, 1966; Sheridan and Rouse, 1971) have found limitations in addition to memory and observation noise in situations where the equivalent of several dots into the future must be predicted.

Pollay's Model of Decision Times for Choices Among Equally Attractive Alternatives

References

Pollay, R. W. "A model of decision times in difficult decision situations," Psychol. Rev., 1970, 77(4), 274-281.

Hendrick, C., Mills, J. and Kiesler, C. A. "Decision time as a function of the number and complexity of equally attractive alternatives," J. Pers. and Soc. Psychol., 1968, 8, 313-318.

Kiesler, C. A. "Conflict and the number of choice alternatives," Psychol. Rep., 1966, 18, 603-610.

Description

This model was developed to predict the apparently anomalous finding of Hendrick, et al (1968) and Kiesler (1966) that decision makers took less time when all four members of a set of alternatives were equally attractive than when two of the members were clearly inferior and easily rejectable. The model requires the following assumptions:

1) The cost of the DM's analytical effort is a linear function of the time spent analyzing.

2) The DM is efficient in assimilating experience and modifying his expectations. If not Bayesian, he is near-Bayesian.

3) There is a subjective utility associated with selection of the "best" alternative that exceeds that alternative's objective utility.

4) Least attractive alternatives are eliminated first.

5) Elimination of alternatives from a decision set follows a Poisson process, with the result that decision times are exponentially distributed.

The DM enters the decision situation with some prior notion of the difficulty of the task with which he is presented. He will continue his analysis of alternatives until the expected utility of the "best" choice is equal to or less than the expected cost of additional analysis, at which time he will make a selection and terminate his considerations.

The mathematical formulation of the model leads to four specific predictions:

1) When it is easy to discriminate between members of a set of alternatives, the difference between the decision time when all members of the set are good and that when two (i.e., half the set) are inferior will be positive.

2) When it is difficult to discriminate between set members, the difference between the decision time when all are good and that when half the set are inferior will be negative.

3) When discrimination is easy, the difference between decision times for all-good and half-good sets, respectively, will be uncorrelated with personality variables (as specified, e.g., by the California Personality Inventory, CPI).

4) When discrimination between good alternatives is difficult, the differences in decision times obtained with sets of all-good alternatives and one-half-good, respectively, will be positively correlated with CPI personality variables such as confidence, self-reliance and persistence (Scales I and III) and uncorrelated with socialization and responsibility variables (Scales II and IV).

Input Parameters

Prior distributions, a measure of the equivalence of alternatives, utilities and costs must be input.

Model Outputs

The model predicts the phenomenon observed by Kiesler and by Hendrick as a function of the difficulty of discrimination between attractive alternatives. In addition, it predicts correlations between the magnitude of the phenomenon and certain personality variables.

Model Validation

No direct validations are reported by the author. However, a set of his observations in another research setting and the findings of other investigators having to do with aspects of the general phenomenon are considered to lend support to the validity of the model.

Comments

If valid, this model may have considerable value in the simulation of decision processes involving multiple alternatives which, either because they are incompletely specified or because they are actually equivalent, in some sense are difficult to disciminate among. Beyond the simulation effort itself, the model might also be used diagnosticaly to uncover equivalence classes that account for unnecessary delays and, hence, lead to more efficient statements of alternative courses of action. The model is also intriguing in the sense that it incorporates a concept of "decision style" and leads to researchable hypotheses concerning personality variables.

Wingert's Function Interlace Model for Workload Prediction

References

Wingert, J. W. "Function interlace modifications to analytic workload prediction," (in Cross, R. D. & McGrath, J. J. eds.) Crew System Design, JANAIR Proc., July 1973.

Lindquist, O. H., Jones, A. L., & Wingert, J. W. An investigation of airborne displays and controls for search and rescue. JANAIR Report 701221, 1971.

Description

This model was developed for the purpose of estimating workload in tasks where the operator appears to be able to share time between continuous and discontinuous subtasks. The author makes the point that workload associated with such tasks may be very much overestimated by models that view the operator as an essentially single-channel device, particularly when he is well-trained and highly experienced.

The basic idea of the Function Interlace Model is expressed by the equation.

where ω_1 and ω_2 = the instantaneous workloads predicted for subtask 1 and subtask 2, respectively; and I = the interlace coefficient associated with the combined subtasks.

To use the model one must first derive estimates of the workloads associated with the subtasks when each is performed independently. By analysis or by simulation techniques employing a tertiary task designed to provide maximum interference with the primary subtasks, the interlace coefficient is then estimated.

Input Parameters

Estimates of workload associated with each subtask and an estimate of the interlace coefficient are required.

Model Output

The model provides an estimate of workload in tasks permitting parallel processing.

Model Validation

The model is still in evaluation and has not yet undergone complete validation. Primary attention has been focussed on validating the techniques (e.g., eye movements) for assessment of workload of independent subtasks.

Comments

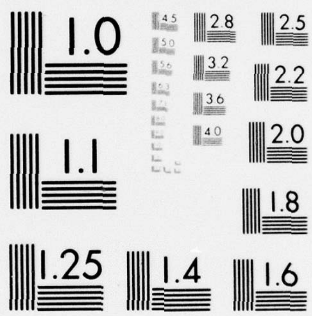
This model appears to require a good deal of artistry during assessment of independent and interactive workload, particularly where simulation is not a viable method for establishing baseline data. Also, there is a question as to how interlace effects can be estimated for any group of subtasks without making the equivalent of a complete factorial analysis in which each subtask is combined with each other. If this need be done, one wonders about the justification for considering "Function Interlace" to be a model. We suspect that a major value in viewing operator workload in this framework is that one is forced to pay very close attention to the structural requirements and the inputs and outputs of the subtask sequence - hence, to become intimately familiar with the demands made upon the operator. Perhaps, with this familiarity, the assessment of workload is quite straightforward and of high reliability.

AD-A038 597 BOLT BERANEK AND NEWMAN INC CAMBRIDGE MASS F/G 5/8
CRITICAL REVIEW AND ANALYSIS OF PERFORMANCE MODELS APPLICABLE T--ETC(U)
MAR 77 R W PEW, C E FEEHRER, S BARON F44620-76-C-0029

UNCLASSIFIED BBN-3446 AFOSR-TR-77-0520 NL

4 OF 4
ADA038597

END
DATE
FILMED
5 - 77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

REFERENCES

REFERENCES

Adkins, L.A., St. Clair, J.D., Thatro, M.C., & Willis, H.R. Development of all-weather landing system reliability analysis and criteria for category III airborne systems (Vol. III). Lockheed-Georgia, FAA/SRDS Report No. RD-67-20, May 1967.

Allen, R.W., Clement, W.F., & Jex, H.R. Research on display scanning, sampling, and reconstruction using separate main and secondary tasks. NASA CR-1569, July 1970.

Altman, J.W. Classification of human error. In W.B. Askren (Ed.), Symposium on Reliability of Human Performance in Work, AMRL-TR-67-88, 1967, 3-16.

Anderson, R.O. A new approach to the specification and evaluation of flying qualities. AFFDL-TR-69-120, Wright-Patterson Air Force Base, June 1970.

Anderson, R.O., Connors, A.J., & Dillow, J.D. Paper pilot ponders pitch. AFFDL/FGC-TM-70-1, Wright-Patterson Air Force Base, January 1971.

Arad, B. Notes on the measurement of control load and sector design in the enroute environment. Appendix to final report on Project 102-11, June 1964.

Arad, B., Golden, B.T., Grambard, J.E., Mayfield, C.E., & van Saun, H.R. Control load, control capacity, and optimal sector design. FAA Report RD-64-16, Project No. 102-11R, December 1963.

Arnold, J.D., Johnson, R.B., & Dillow, J.D. Pitch paper pilot revisited. Proceedings of Ninth Annual Conference on Manual Control, M.I.T., Cambridge, Mass., May 1973.

Atkinson, R.C., & Shiffrin, R.M. Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.), The psychology of learning and motivation, Vol. 2, New York: Academic Press, 1968, 89-195.

Audley, R.J., & Pike, A.R. Some alternative stochastic models of choice. British Journal of Mathematical and Statistical Psychology, 1965, 18, 207-225.

Barnes, J.A. Analysis of pilot's eye movements during helicopter flight. U.S. Army Aberdeen Research and Development Center TN-11-72, April 1972.

Baron, S., Kleinman, D.L., Miller, D.C., Levison, W.H., & Elkind, J.I. Application of optimal control theory to the prediction of human performance in a complex task. Wright-Patterson Air Force Base, AFFDL-TR-69-81, March 1970.

Baron, S., & Kleinman, D.L. The human as an optimal controller and information processor. NASA CR-1151, September 1968.

Baron, S., & Levison, W.H. Analysis and modelling human performance in AAA tracking. BBN Report No. 2557, Bolt Beranek and Newman Inc., Cambridge, Mass., March 1974.

Bekey, G.A. The human operator as a sampled-data system. IRE Transactions on Human Factors in Electronics, 1962, HFE-3, 43-51.

Blanchard, R.E. Requirements, concept, and specification for a Navy human performance data store. Behaviormetrics Report No. 102-2, April 1973.

Brainard, R.W., et al. Some variables influencing the rate of gain of information. Journal of Experimental Psychology, 1962, 63, 105-110.

Broadbent, D.E. Decision and stress. London: Academic Press, 1971.

Broadbent, D.E. Perception and communication. London: Pergamon Press, 1958.

Buckner, D., & McGrath, J. A comparison of performance on single and dual mode vigilance tasks: Human Factor Problems in Anti-Submarine Warfare Technical Report No. 8, Human Factors Research Inc., 1961.

Bush, R.R., & Mosteller, F. Stochastic models for learning. New York: J. Wiley and Sons, 1955.

Carbonell, J.R. A queueing model of many-instrument visual sampling. IEEE Transactions on Human Factors in Electronics, December 1966, HFE-7(4), 157-163.

Chernikoff, R., et al. Two dimensional tracking with identical and different control dynamics in each coordinate. NRL Report 5424, November 1959.

Clark, C., & Williges, R. Response surface methodology design variants useful in human performance research. University of Illinois, Aviation Psychology Laboratory Report No. ARL-71-31/AFOSR-71-7, 1971.

Clark, C., & Williges, R.C. Response surface methodology central-composite design modifications for human performance research. Human Factors, 1973, 15(4), 295-309.

Cochran, D., Purswell, J., & Hoag, L. Development of a prediction model for dynamic visual inspection tasks. Proceedings of the Human Factors Society, 1973, 31-51.

Connolly, D.W., & McCosker, W.R. Human factors in use of terminal radar (analogue) display systems. U.S. Federal Aviation Administration, National Aviation Facilities Experimental Center, Report No. FAA-RD-70-66, 1970.

Costello, R.G. The surge model of the well-trained human operator in simple manual control. IEEE Transactions on Man-Machine Systems, 1968, MMS-9, 2-9.

David, H. Maneuvers recommended by air traffic controllers observing electronic simulations and their estimated consequences. Loughborough University of Technology, Department of Ergonomics and Cybernetics, 1967.

Diamantides, N. A pilot analog for airplane pitch control. Journal of Aeronautical Science, 1958, 25, 361-371.

Dillow, J.D. The "paper pilot" -- A digital computer program to predict pilot rating for the hover task. Wright-Patterson Air Force Base, AFFDL-TR-70-40, March 1971.

Dodson, J.D. Simulation system design for a TEAS simulation research facility. Report AFCRL 1112 and PRC 194. Planning Research Corporation, Los Angeles, November 15, 1961.

Donders, F.C. On the speed of mental processes. In W.G. Koster (Ed. and Trans.), Attention and performance II. Amsterdam: North-Holland, 1969. Reprinted from Acta Psychologica, 1969,

30.

Drury, C. Application of Fitts' law to foot pedal design. Human Factors, 1975, 17, 368-373.

Drury, C.G. Inspection of sheet materials - model and data. Human Factors 1975, 17(3), 257-265.

Drury, C.G. The effect of speed of working on industrial inspection accuracy. Applied Ergonomics, 1973, 4(1), 2-7.

Dunlay, W.J., Horonjeff, R., & Kanafani, A. Models for estimating the number of conflicts perceived by air traffic controllers. Institute of Transportation and Traffic Engineering, University of California, Berkeley, Department of Transportation Project: FA-72-WA-2827, December 1973.

Edwards, W. Behavioral decision theory. Annual Review of Psychology, 1961, 12, 473-493.

Edwards, W. Optimal strategies for seeking information: Models for statistics, choice reaction times and human information processing. Journal of Mathematical Psychology, 1965, 2, 312-329.

Edwards, W., Phillips, L.D., Hays, W.L., & Goodman, B.C. Probabilistic information processing systems: Design and evaluation. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 248-265.

Egan, J.P., Greenberg, G.Z., & Schulman, A.L. Operating characteristics, signal detectability and the method of free response. Journal of Acoustical Society of America, 1961, 33, 993-1007.

Elkind, J.I., & Miller, D.C. Adaptive characteristics of the human controller of time-varying systems. BBN Report No. 1360, Bolt Beranek and Newman Inc., Cambridge, Mass., April 1966; also AFFDL-TR-66-60, Wright-Patterson Air Force Base, Ohio, December 1967.

Elkind, J.I. Characteristics of simple manual control systems. M.I.T. Lincoln Laboratory TR-111, April 1956.

Elkind, J.I. A survey of the development of models for the human controller. In R.c. Langford & C.J. Mundo (Eds.), Guidance and Control - II, Vol. 13 of Progress in Astronautics and Aeronautics. New York: Academic Press, 1964.

Feigenbaum, E. The simulation of verbal learning behavior. Proceedings of the Western Joint Computer Conference, 1961, 19, 121-132.

Fogel, L.J., & Moore, R.A. Modelling the human operator with finite-state machines. Proceedings of Fourth Annual NASA - University Conference on Manual Control, NASA SP-192, 1968, 509-526.

Fox, J.G., & Haselgrove, C.M. Industrial inspection efficiency and the probability of a defect occurring. Ergonomics, 1969, 12(5), 713-721.

Funaro, J.F. An empirical analysis of five descriptive models for cascaded inference. Rice University, Department of Psychology Research Report Series, Report No. 74-2, May 1974.

Fussell, J.B., Powers, G.J., & Bennetts, R.G. Fault trees - A state of the art discussion. IEEE Transactions on Reliability, 1974, 23(1), 51-53.

Gai, E.G., & Curry, R.E. A model of the human observer in failure detection tasks. IEEE Transactions on Systems, Man and Cybernetics, 1976, SMC-6, 85-94.

Gettys, C.F., & Wilke, T.A. The application of Bayes's theorem when the true data state is uncertain. Organizational Behavior and Human Performance, 1969, 4, 125-141.

Green, D.M., & Swets, J.A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Greening, C.P. Mathematical modelling of air-to-ground target acquisition. Human Factors, 1976, 18(2), 11-148.

Groen, G.J., & Parkman, J.M. A chronometric analysis of simple addition. Psychological Review, 1972, 79, 329-343.

Hann, R., & Kuperman, G. SAINT model of a choice reaction time paradigm. Proceedings of 19th Annual Meeting of the Human

Factors Society, 1975, 336-341.

Harper, R.P. Jr., & Cooper, G.E. A revised pilot rating scale for the evaluation of handling qualities. Cornell Aero. Labs, Report 153, September 1966.

Hendrick, C., Mills, J., & Kiesler, C.A. Decision time as a function of the number and complexity of equally attractive alternatives. Journal of Personality and Social Psychology, 1968, 8, 313-318.

Hoffman, L.G., Clement, N.F., & Blodgett, R.E. Further examination of pilot instrument scanning data and development of a new link value estimator. Proceedings of the Ninth Annual Conference on Manual Control, M.I.T., Cambridge, Mass., May 23-25, 1973.

Irwin, I.A., Levitz, J.J., & Freed, A.M. Human reliability in the performance of maintenance. Proceedings of the Symposium on Quantification of Human Performance Albuquerque, N.M., 1964.

Jex, H.R., Allen, R.W., & Magdaleno, R.E. Display format effects on precision tracking performance, describing functions and remnant. U.S. Air Force Aerospace Medical Division, AMRL-TR-71-63, August 1971.

Johannsen, G. Development and optimization of a nonlinear multiparameter human operator model. IEEE Transactions on Systems, Man, and Cybernetics, 1972, SMC-2(4), 494-503.

Kahneman, D. Attention and effort. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1973.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.

Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

Karger, D.W., & Bayha, F.H. Engineered work measurement. New York: Industrial Press, 1966.

Kiesler, C.A. Conflict and the number of choice alternatives. Psychological Reports, 1966, 18, 603-610.

Kleinman, D.L., Baron, S., & Levison, W.H. An optimal-control model of human response, Part 1: Theory and validation. Automatica, 1970, 6, 357-369.

Kleinman, D.L., & Baron, S. A control theoretic model for piloted approach to landing. Automatica, 1973, 9, 339-347.

Kleinman, D.L., & Killingsworth, W.R. A predictive pilot model for STOL aircraft landing. NASA CR-2374, March 1974.

Kleinman, D.L., & Perkins, T. Modeling the human in a time-varying anti-aircraft tracking loop. IEEE Transactions on Automatic Control, 1974, AC-19, 297-306.

Knoop, D.E., & Fu, K.S. An adaptive model of the human operator in a control system. Proceedings of Fifth National Symposium on Human Factors in Electronics, May 1964.

Knowles, W.B., et al. Models, measures and judgments in system design. Human Factors, 1969, 11(6), 577-590.

Kuperman, G., & Seifert, D. Development of a computer simulation model for evaluating DAIS display concepts. Proceedings of 19th Annual Meeting of the Human Factors Society, 1975, 347-353.

Kvalseth, T.O. A model of linear arm movements with preview constraints. Ergonomics, 1975, 18, 529-538.

Lamb, J.C. Reliability and independence. In J.P. Jenkins (Ed.), Proceedings of U.S. Navy Human Reliability Workshop, 1970, 208-215.

Laming, D.R.J. Information theory of choice-reaction times. London: Academic Press, 1968.

Leonard, T. Optimizing linear dynamics for human-operated systems by minimizing the mean-squared tracking error. WESCON, 1960, Vol. 4, Part 4.

Levison, W.H., Baron, S., & Junker, A.M. Modeling the effects of environmental factors on human control and information

processing. AMRL-TR-76-74, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, August 1976.

Levison, W.H., Baron, S., & Kleinman, D.L. A model for human controller remnant. IEEE Transactions on Man-Machine Systems, 1969, MMS-10(4), 101-107.

Levison, W.H., Elkind, J.I., & Ward, J.L. Studies of multivariable manual control systems: A model for task interference. NASA-Ames Research Center, NASA CR-1746, May 1971.

Levison, W.H., & Elkind, J.I. Studies of multivariable manual control systems: Two-axis compensatory systems with separated displays and controls. NASA CR-875, October 1967.

Levison, W.H., & Houck, P.D. Guide for the design of control sticks in vibration environments. Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, AMRL-TR-74-127, February 1975.

Levison, W.H., & Tanner, R.B. A control-theory model for human decision making. NASA CR-1953, December 1971.

Lindquist, O.H., Jones, A.L., & Wingert, J.W. An investigation of airborne displays and controls for search and rescue. JANAIR Report 701221, 1971.

Lucas, P.A. Human performance in low signal probability tasks. Journal of Acoustical Society of America, 1967, 42, 158-178.

Luce, R.D., & Green, D.M. Detection of auditory signals presented at random times. Perception and Psychophysics, 1970, 7, 1-14.

Lusted, L.B. Introduction to medical decision making. Springfield: Charles C. Thomas, 1968.

MacCrimmon, K., & Ryavec, C. An analytical study of the PERT assumptions. Operations Research, 1964, 12, 16-37.

Machuca, L.A., & Lind, J.M. Verification and extension of display scanning and remnant models using aircraft lateral and longitudinal dynamics as controlled elements. Air Force

Institute of Technology, MS Thesis, GGC/EE/71-14, May 1971.

Maltas, K.L., & Buck, J.R. A man-machine system study of manual and automated hot strip mills. Purdue Laboratory for Applications Indus. Ctrl. Report No. 66, March 1975.

Mangelsdorf, J.E. Variables affecting the accuracy of collision judgments on radar-type displays. Wright Air Development Center, Report No. WADC-TR-55-462, 1955.

Martin, W.L. Stress considerations in man-machine simulation. 4th Annual Psychology in the Air Force Symposium Proceedings, USAF Academy, Colorado. April 1974. (AMRL-TR-74-34).

McCornack, R.L. Inspector accuracy: A study of the literature. SANDIA Report No. SCTM 53-61(14), February 1961.

McRuer, D.T., Graham, D., Krendel, E.S., & Reisener, W. Human pilot dynamics in compensatory systems. Wright-Patterson Air Force Base, AFFDL-TR-65-15, July 1965.

McRuer, D.T., Krendel, E.S. Dynamic response of human operators. Wright Air Development Center, WADC TR-56-524, October 1957.

McRuer, D.T., & Krendel, E.S. Mathematical models of human pilot behavior. AGARD Report AGARD-AG-188, January 1974.

McRuer, et al. A systems analysis theory for displays in manual control. Systems Technology Inc. Report No. 163-1, June 1968.

Meister, D. Comparative analysis of human reliability models. Bunker Ramo, Report No. L0074-107, November 1971.

Meister, D. Methods of predicting human reliability in man-machine systems. Human Factors, 1964, 6(6), 621-646.

Meister, D., & Mills, R.G. Development of a human performance reliability data system: Phase I. Report No. F33615-70-C-1518, Bunker-Ramo, Cal., 1971.

Melton, A.W. Implications of short term memory for a general theory of memory. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 1-21.

Miller, D.C. A model for the adaptive response of the human controller to sudden changes in controlled process dynamics. S.M. Thesis, M.I.T., June 1965.

Miller, D.P., & Vinje, E.W. Fixed-base flight simulator studies of VTOL aircraft handling qualities in hovering and low speed flight. Wright-Patterson Air Force Base, Ohio, AFFDL-TR-67-152, January 1968.

Miller, R.W. Schedule, cost and profit control with PERT. New York: McGraw Hill, 1963.

Mills, R. Some recent efforts toward development of a human performance reliability data system and supportive human performance reliability principles. In J. Jenkins (Ed.), Proceedings of U.S. Navy Human Reliability Workshop, 1971.

Mills, R.G. System modeling and simulation: Application of a research methodology and test. Unpublished Ph.D. Dissertation, Ohio State University, 1976.

Mills, R.G., Bachert, R.F., & Aume, N.M. Summary report of AMRL remotely piloted vehicle (RPV) system simulation study: II results. Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, Ohio, AMRL-TR-75-13, 1975.

Mills, R.G., & Hatfield, S.A. Sequential task performance: Task modules relationships, reliabilities and times. Human Factors, 1974, 16(2), 117-133.

Morton, J. A functional model for memory. In D.A. Norman (Ed.), Models for human memory. New York: Academic Press, 1970.

Newell, A., Shaw, J.C., & Simon, H.A. Elements of a theory of human problem solving. Psychological Review, 1958, 65, 151-166.

Newell, A., & Simon, H.A. GPS, a program that simulates human thought. In E. Feigenbaum & J. Feldman (Eds.), Computers and thought. New York: McGraw Hill, 1963.

Nickerson, R.S., & Feehrer, C.E. Decision making and training: A review of theoretical and empirical studies of decision making and their implications for the training of decision making.

NAVTRAEQUIPCEN 73-C-0128-1, July 1975.

Niemala, R.J., & Krendel, E.S. Detection of a change in plant dynamics in a man-machine system. Proceedings of Tenth Annual NASA-University Conference on Manual Control, Wright-Patterson Air Force Base, Ohio, April 9-11, 1974.

Norman, D.A., & Rumelhart, D.E. A system for perception and memory. In D.A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970.

Obermayer, R.W., & Muckler, F.A. On the inverse optimal control problem in manual control systems. NASA CR-208, 1965.

Onstott, E.D. Multi-axis pilot-vehicle dynamics. Proceedings of the Tenth Annual NASA-University Conference on Manual Control, Wright-Patterson Air Force Base, Ohio, 1974.

Payne, D., & Altman, J.W. An index of electronic equipment operability: Report of development. American Institutes for Research, January 1962.

Peterson, C.R., & Beach, L.R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68(1), 29-46.

Phatak, A., Weinert, H., & Segall, I. Identification of the optimal control model for the human operator. Wright-Patterson Air Force Base, AMRL-TR-74-79, September 1974.

Phatak, A., & Kleinman, D.L. Current status of models for the human operator as a controller and decision maker in manned aerospace systems. Automation in manned aerospace systems. AGARD Proceedings No. 114, Dayton, Ohio, October 1972.

Phatak, A.U., & Bekey, G.A. Model of the adaptive behavior of the human operator in response to a sudden change in the control situation. IEEE Transactions on Man-Machine Systems, 1969, MMS-10, 72-80.

Phillips, L.D., & Edwards, W. Conservation in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-357.

Pitz, G.F. Sample size, likelihood and confidence in a decision. Psychonomic Science, 1967, 8(6), 257-258.

Pollay, R.W. A model of decision times in difficult decision situations. Psychological Review, 1970, 77(4), 274-281.

Posner, M.I. Information reduction in the analysis of sequential tasks. Psychological Review, 1964, 71, 491-504.

Preyss, A.E. A theory and model of human learning behavior in a manual control task. NASA CR-1124, August 1968.

Pritsker, A.A.B., Wortman, D.B., Seum, C.S., Chubb, G.P., & Seifert, D.J. SAINT: Volume I. Systems analysis of an integrated network of tasks. AMRL-TR-73-126, April 1974.

Pritsker, A.A.B., & Happ, W.W. GERT: Graphical evaluation and review technique. Part I. Fundamentals. Journal of Industrial Engineering, 1967, 17, 267-274.

Ratner, R.S., Williams, J.O., Glaser, M.B., Stuntz, S.E., & Couluris, G.J. The air traffic controller's contribution to ATC system capacity in manual and automated environments. FAA Report RD-72-63, Volume II, June 1972.

Restle, F., & Greeno, J.G. Introduction to mathematical psychology. Reading, Ma.: Addison-Wesley Publishing Co., 1970.

Rigby, L.V. The Sandia human-error rate bank (SHERB). Human Factors Soc. - L.A. Chap., 1967, pp. (5-1) - (5-13).

Roig, R.W. A comparison between human operator and optimum linear controller RMS-error performance. IRE Transactions on Human Factors Electronics, 1962, HFE-3, 18-22.

Rouse, W.B. A model of the human in a cognitive prediction task. IEEE Transactions on Systems, Man and Cybernetics, 1973, SMC-3(5), 473-477.

Rumelhart, D.E. A multicomponent theory of the perception of briefly exposed visual displays. Journal of Mathematical Psychology, 1970, 7, 191-218.

Rumelhart, D.E., & Siple, P. Process of recognizing tochistoscopically presented words. Psychological Review, 1974, 81, 99-118.

Russell, L. Characteristics of the human as a linear servo-element. M.I.T., MS Thesis, May 1951.

Saunders, R.S., Smith, M.G., & Teichner, W.H. Models of short term memory: A critical review. Department of Psychology, New Mexico State University Technical Report 74-1, February 1974.

Schum, D.A., & DuCharme, W.M. Comments on the relationship between the impact and the reliability of evidence. Organizational Behavior and Human Performance, 1971, 6, 111-131.

Senders, J.W. The human operator as a monitor and controller of multi-degree of freedom systems. IEEE Transactions on Human Factors in Electronics, 1964, HFE-5, 2-5.

Senders, J.W., Elkind, J.I., Grignetti, M.C., & Smallwood, R. An investigation of the visual sampling behavior of human observers, NASA CR-434, April 1966.

Senders, J.W., Ward, J.L., & Carbonell, J.R. Human visual sampling processes: A simulation validation study. NASA CR-1258, January 1969.

Sheehan, J.J., & Drury, C.G. The analysis of industrial inspection. Applied Ergonomics, 1971, 2(2), 74-78.

Sheridan, T.B. On how often the supervisor should sample. IEEE International Symposium on Man-Machine Systems, IEEE Conference Record No. 69C58-MMS, 1969, Vol. 3.

Siegel, A.I., & Wolf, J.J. A technique for evaluating man-machine system designs. Human Factors, 1961, 3(1), 18-28.

Siegel, A.I., & Wolf, J.J. Man-machine simulation models. New York: John Wiley & Sons, Inc., 1969.

Simon, H.A. Models of man. New York: John Wiley & Sons, 1957.

Smallwood, R. Internal models and the human instrument monitor. Symposium on Human Factors in Electronics, IEEE, Minneapolis,

Minnesota, May 1966.

Snapper, K.J., & Fryback, D.G. Inference based on unreliable reports. Journal of Experimental Psychology, 1971, 87, 401-404.

Snapper, K.J., & Peterson, C.R. Information seeking and data diagnosticity. Journal of Experimental Psychology, 1971, 87, 429-433.

Stapleford, R.L., Craig, S.J., & Tennant, J.A. Measurement of pilot describing functions in single-controller multiloop tasks. NASA CR-1238, January 1969.

Stapleford, R.L., McRuer, D.T., & Magdaleno, R.E. Pilot describing function measurements in a multiloop task. IEEE Transactions on Human Factors in Electronics, 1967, HFE-8(2)

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and performance II. Amsterdam: North-Holland, 1969. Reprinted from Acta Psycholgica, 1969, 30.

Stone, J.R. Paper pilot ponders supersonic transports. Proceedings of the Ninth Annual Conference on Manual Control, M.I.T., Cambridge, Mass., May 1973.

Strieb, M.I. The human operator simulator. Volume I Introduction and Overview. Analytics Technical Report 1117-1, August 1975.

Swain, A. Development of a human error rate data bank. In J. Jenkins (Ed.), Proceedings U.S. Navy Human Reliability Workshop, 1971.

Swain, A.D. Shortcuts in human reliability analysis. Paper presented at NATO Advanced Study Institute on Generic Techniques in Systems Reliability Accessment, University of Liverpool, England, July 16-27, 1973.

Swain, A.D. Some limitations in using the simple multiplicative model in behavior quantification. In W.B. Askren (Ed.), Symposium on Reliability of Human Performance in Work, AMRL-TR-67-88, 1967, 17-31.

Swain, A.D. The human element in system development. Proceedings of 1970 Annual Symposium on Reliability, IEEE, 1970, 20-28.

Swain, A.D. Human factors associated with prescribed action links. SANDIA Corp. SAND74-0051, 1974.

Swain, A.D. An error-cause removal program for industry. Human Factors, 1973, 15, 129-136.

Swain, A.D. THERP. Albuquerque: Sandia Corp., SCR-64-1338, 1964.

Swain, A.D., & Guttman, H.E. Human reliability analysis applied to nuclear power. Proceedings of 14th Annual Reliability and Maintainability Symposium, IEEE, January 1975, 116-119.

Swets, J.A. Monitoring and attention sharing. Chapter 2 in the C3 System User Vol. I: A review of research on human performance as it relates to the design and operation of command, control and comunication systems. Bolt Beranek and Newman Inc. Report No. 3459, February 1977.

Tanner, W.P., & Swets, J.A. A decision making theory of visual detection. Psychological Review, 1954, 61, 401-409.

Teichner, W.H. Quantitative models for predicting human visual/perceptual/motor performance. Department of Psychology, New Mexico State University Technical Report 74-3, October 1974.

Theologus, G.C., & Fleishman, E.A. Development of a taxonomy of human performance: Validation study of ability scales for classifying human tasks. American Institutes for Research Technical Report 10, April 1971.

Thomas, H.G.B. Communication theory and the constellation hypothesis of calculation. Quarterly Journal of Experimental Psychology, 1963, 15, 173-191.

Thomas, M.U. A human response model of a combined manual and decision task. IEEE Transactions on Systems, Man and Cybernetics, 1973, SMC-3(5), 478-484.

Thomas, M.U. Some probabilistic aspects of performance times in a combined manual and decision task. Ph.D. dissertation, University of Michigan, Ann Arbor, 1971.

Todosiev, E.P. et al. Human tracking performance in uncoupled and coupled two-axis systems. TRW Systems, 4380-6003-R-0000, December 1965.

Tustin, A. An investigation of the operator's response in manual control of a power driven gun. Metropolitan-Vickers Electrical Co., Ltd., Attercliffe Common Works, Sheffield, England. C.S. Memorandum No. 169, August 1944.

Tustin, A. The nature of the operator's response in manual control and its implications for controller design. Journal of IEE, 1947, 94, Part IIA, No. 2.

Tversky, A., & Kahneman, D. Belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.

Tversky A., & Kahneman, D. Judgment under uncertainty: heuristics and biases. Science, 1974, 185, 1124-1131.

Veldhuyzen, W. Ship manoeuvring under human control. Doctoral dissertation, Delft University, The Netherlands, June 1976.

Verdi, A.P., & Smith, R.L. Personnel performance data store for navy systems: I. Preliminary program concept. Dunlap and Assoc., Inc., Santa Monica, Ca., July 1966.

Vesely, W.E., & Narum, R.E. PREP and KITT: Computer codes for automatic evaluation of a fault-tree. IN-1349, August 1970. (Available through NTIS, Springfield, VA 22151.)

Vickers, D. Evidence for an accumulator model of psychophysical discrimination. Ergonomics, 1970, 13, 37-58.

Wallack, P.M., & Adams, S.K. The utility of signal detection theory in the analysis of industrial inspection accuracy. AIIE Transactions, 1969, 1(1), 33-44.

Weir, D.H., & McRuer, D.T. Pilot dynamics for instrument approach tasks: Full panel multiloop and flight director operations. NASA CR-2019, May 1972.

Wewerinke, P.H. Human operator workload for various control situations. Tenth Annual Conference on Manual Control, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 1974.

Williges, R., & Mills, R. Predictive validity of central composite design regression equations. Human Factors, 1973, 15(4), 349-354.

Wingert, J.W. Function interlace modifications to analytic workload prediction. In R.D. Cross & J.J. McGrath (Eds.), Crew system design, JANAIR Proceedings, July 1973.

Wortman, D., Duket, S., & Seifert, D. SAINT simulation of a remotely piloted vehicle/drone control facility. Proceedings of 19th Annual Meeting of the Human Factors Society, 1975, 342-346.

Wortman, D.B., Pritsker, A.A.B., Seum, C.S., Seifert, D.J., & Chubb, G.P. SAINT Volume II User's Manual. AMRL-TR-73-128, Aerospace Medical Research Laboratory, Wright Patterson AFB, Ohio, 1974.

Young , L.R. On adaptive manual control. Ergonomics, 1969, 4, 635.

Young L.R. Human control capabilities. In J.F. Parker, Jr. & V.R. West (Eds.), Bioastronautics data book, Chapter 16, 2nd edition, NASA SP-3006, 1973, 751-806.

Young L.R., & Stark, L. Biological control systems - A critical review and evaluation. NASA CR-190, 1965.